(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 786 519 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
30.07.1997 Bulletin 1997/31

(51) Int Cl.[6]: **C12N 15/00**

(21) Application number: **97100117.7**

(22) Date of filing: **07.01.1997**

(84) Designated Contracting States:
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV RO SI

(30) Priority: **05.01.1996 US 9861**

(71) Applicant: **HUMAN GENOME SCIENCES, INC.**
Rockville, MD 20850-3338 (US)

(72) Inventors:
- **Kunsch, Charles A.**
  Gaithersburg, Maryland 20882 (US)
- **Choi, Gil H.**
  Rockville, Maryland 20850 (US)
- **Barash, Steven C.**
  Rockville, Maryland 20850 (US)
- **Dillon, Patrick J.**
  Gaithersburg, Maryland 20879 (US)
- **Fannon, Michael R.**
  Silver Spring, Maryland 20906 (US)
- **Rosen, Craig A.**
  Laytonsville, Maryland 20882 (US)

(74) Representative: **VOSSIUS & PARTNER**
Postfach 86 07 67
81634 München (DE)

(54) Staphylococcus aureus polynucleotides and sequences

(57) The present invention provides polynucleotide sequences of the genome of *Staphylococcus aureus,* polypeptide sequences encoded by the polynucleotide sequences, corresponding polynucleotides and polypeptides, vectors and hosts comprising the polynucleotides, and assays and other uses thereof. The present invention further provides polynucleotide and polypeptide sequence information stored on computer readable media, and computer-based systems and methods which facilitate its use.

EP 0 786 519 A2

Printed by Jouve, 75001 PARIS (FR)

Description

The present invention relates to the field of molecular biology. In particular, it relates to, among other things, nucleotide sequences of *Staphylococcus aureus,* contigs, ORFs, fragments, probes, primers and related polynucleotides thereof, peptides and polypeptides encoded by the sequences, and uses of the polynucleotides and sequences thereof, such as in fermentation, polypeptide production, assays and pharmaceutical development, among others.

The genus *Staphylococcus* includes at least 20 distinct species. (For a review see Novick, R. P., The *Staphylococcus* as a Molecular Genetic System, Chapter 1, pgs. 1-37 in MOLECULAR BIOLOGY OF THE *STAPHYLOCOCCI,* R. Novick, Ed., VCH Publishers, New York (1990)). Species differ from one another by 80% or more, by hybridization kinetics, whereas strains within a species are at least 90% identical by the same measure.

The species *Staphylococcus aureus,* a gram-positive, facultatively aerobic, clump-forming cocci, is among the most important etiological agents of bacterial infection in humans, as discussed briefly below.

Human Health and *S. Aureus*

Staphylococcus aureus is a ubiquitous pathogen. (See, for instance, Mims *et al.,* MEDICAL MICROBIOLOGY, Mosby-Year Book Europe Limited, London, UK (1993)). It is an etiological agent of a variety of conditions, ranging in severity from mild to fatal. A few of the more common conditions caused by *S. aureus* infection are burns, cellulitis, eyelid infections, food poisoning, joint infections, neonatal conjunctivitis,osteomyelitis, skin infections, surgical wound infection, scalded skin syndrome and toxic shock syndrome, some of which are described further below.

Burns

Burn wounds generally are sterile initially. However, they generally compromise physical and immune barriers to infection, cause loss of fluid and electrolytes and result in local or general physiological dysfunction. After cooling, contact with viable bacteria results in mixed colonization at the injury site. Infection may be restricted to the non-viable debris on the burn surface ("eschar"), it may progress into full skin infection and invade viable tissue below the eschar and it may reach below the skin, enter the lymphatic and blood circulation and develop into septicaemia. *S. aureus* is among the most important pathogens typically found in burn wound infections. It can destroy granulation tissue and produce severe septicaemia.

Cellulitis

Cellulitis, an acute infection of the skin that expands from a typically superficial origin to spread below the cutaneous layer, most commonly is caused by *S. aureus* in conjunction with *S. pyrogenes.* Cellulitis can lead to systemic infection. In fact, cellulitis can be one aspect of synergistic bacterial gangrene. This condition typically is caused by a mixture of *S. aureus* and microaerophilic *streptococci.* It causes necrosis and treatment is limited to excision of the necrotic tissue. The condition often is fatal.

Eyelid infections

S. aureus is the cause of styes and of sticky eye" in neonates, among other eye infections. Typically such infections are limited to the surface of the eye, and may occasionally penetrate the surface with more severe consequences.

Food poisoning

Some strains of *S. aureus* produce one or more of five serologically distinct, heat and acid stable enterotoxins that are not destroyed by digestive process of the stomach and small intestine (enterotoxins A-E). Ingestion of the toxin, in sufficient quantities, typically results in severe vomiting, but not diarrhoea. The effect does not require viable bacteria. Although the toxins are known, their mechanism of action is not understood.

Joint infections

S. aureus infects bone joints causing diseases such osteomyelitis.

Osteomyelitis

S. aureus is the most common causative agent of haematogenous osteomyelitis. The disease tends to occur in

children and adolescents more than adults and it is associated with non-penetrating injuries to bones. Infection typically occurs in the long end of growing bone, hence its occurrence in physically immature populations. Most often, infection is localized in the vicinity of sprouting capillary loops adjacent to epiphysial growth plates in the end of long, growing bones.

Skin infections

S. aureus is the most common pathogen of such minor skin infections as abscesses and boils. Such infections often are resolved by normal host response mechanisms, but they also can develop into severe internal infections. Recurrent infections of the nasal passages plague nasal carriers of *S. aureus*.

Surgical Wound Infections

Surgical wounds often penetrate far into the body. Infection of such wound thus poses a grave risk to the patient. S. aureus is the most important causative agent of infections in surgical wounds. *S. aureus* is unusually adept at invading surgical wounds; sutured wounds can be infected by far fewer *S. aureus* cells then are necessary to cause infection in normal skin. Invasion of surgical wound can lead to severe *S. aureus* septicaemia. Invasion of the blood stream by *S. aureus* can lead to seeding and infection of internal organs, particularly heart valves and bone, causing systemic diseases, such as endocarditis and osteomyelitis.

Scalded Skin Syndrome

S. aureus is responsible for "scalded skin syndrome" (also called toxic epidermal necrosis, Ritter's disease and Lyell's disease). This diseases occurs in older children, typically in outbreaks caused by flowering of *S. aureus* strains produce exfoliation(also called scalded skin syndrome toxin). Although the bacteria initially may infect only a minor lesion, the toxin destroys intercellular connections, spreads epidermal layers and allows the infection to penetrate the outer layer of the skin, producing the desquamation that typifies the diseases. Shedding of the outer layer of skin generally reveals normal skin below, but fluid lost in the process can produce severe injury in young children if it is not treated properly.

Toxic Shock Syndrome

Toxic shock syndrome is caused by strains of *S. aureus* that produce the so-called toxic shock syndrome toxin. The disease can be caused by *S. aureus* infection at any site, but it is too often erroneously viewed exclusively as a disease solely of women who use tampons. The disease involves toxaemia and septicaemia, and can be fatal.

Nocosomial Infections

In the 1984 National Nocosomial Infection Surveillance Study ("NNIS") *S. aureus* was the most prevalent agent of surgical wound infections in many hospital services, including medicine, surgery, obstetrics, pediatrics and newborns.

Resistance to drugs of S. aureus strains

Prior to the introduction of penicillin the prognosis for patients seriously infected with *S. aureus* was unfavorable. Following the introduction of penicillin in the early 1940s even the worst *S. aureus* infections generally could be treated successfully. The emergence of penicillin-resistant strains of *S. aureus* did not take long, however. Most strains of *S. aureus* encountered in hospital infections today do not respond to penicillin; although, fortunately, this is not the case for *S. aureus* encountered in community infections.

It is well known now that penicillin-resistant strains of *S. aureus* produce a lactamase which converts penicillin to pencillinoic acid, and thereby destroys antibiotic activity. Furthermore, the lactamase gene often is propagated episomally, typically on a plasmid, and often is only one of several genes on an episomal element that, together, confer multidrug resistance.

Methicillins, introduced in the 1960s, largely overcame the problem of penicillin resistance in *S. aureus*. These compounds conserve the portions of penicillin responsible for antibiotic activity and modify or alter other portions that make penicillin a good substrate for inactivating lactamases. However, methicillin resistance has emerged in *S. aureus*, along with resistance to many other antibiotics effective against this organism, including aminoglycosides, tetracycline, chloramphenicol, macrolides and lincosamides. In fact, methicillin-resistant strains of *S. aureus* generally are multiply drug resistant.

The molecular genetics of most types of drug resistance in *S. aureus* has been elucidated (See Lyon *et al., Microbiology Reviews* 51: 88-134 (1987)). Generally, resistance is mediated by plasmids, as noted above regarding penicillin resistance; however, several stable forms of drug resistance have been observed that apparently involve integration of a resistance element into the *S. aureus* genome itself.

Thus far each new antibiotic gives rise to resistance strains, stains emerge that are resistance to multiple drugs and increasingly persistent forms of resistance begin to emerge. Drug resistance of *S. aureus* infections already poses significant treatment difficulties, which are likely to get much worse unless new therapeutic agents are developed.

Molecular Genetics *of Staphylococcus Aureus*

Despite its importance in, among other things, human disease, relatively little is known about the genome of this organism.

Most genetic studies of *S. aureus* have been carried out using the the strain NCTC8325, which contains prophages psi11 psi12 and psi13, and the UV-cured derivative of this strain, 8325-4 (also referred to as RN450), which is free of the prophages.

These studies revealed that the *S. aureus* genome, like that of other *staphylococci,* consists of one circular, covalently closed, double-stranded DNA and a collection of so-called variable accessory genetic elements, such as prophages, plasmids, transposons and the like.

Physical characterization of the genome has not been carried out in any detail. Pattee *et al.* published a low resolution and incomplete genetic and physical map of the chromosome of *S. aureus* strain NCTC 8325. (Pattee *et al.* Genetic and Physical Mapping of Chromosome of *Staphylococcus aureus* NCTC 8325, Chapter 11, pgs. 163-169 in. MOLECULAR BIOLOGY OF THE *STAPHYLOCOCCI,* R.P. Novick, Ed., VCH Publishers, New York, (1990) The genetic map largely was produced by mapping insertions of Tn551 and Tn4001, which, respectively, confer erythromycin and gentamicin resistance, and by analysis of SmaI-digested DNA by Pulsed Field Gel Electrophoresis ("PFGE").

The map was of low resolution; even estimating the physical size of the genome was difficult, according to the investigators. The size of the largest SmaI chromosome fragment, for instance, was too large for accurate sizing by PFGE. To estimate its size, additional restriction sites had to be introduced into the chromosome using a transposon containing a SmaI recognition sequence.

In sum, most physical characteristics and almost all of the genes of *Staphylococcus aureus* are unknown. Among the few genes that have been identified, most have not been physically mapped or characterized in detail. Only a very few genes of this organism have been sequenced. (See, for instance Thornsberry, J. , *Antimicrobial Chemotherapy* 21 Suppl C: 9-16 (1988), current versions of GENBANK and other nucleic acid databases, and references that relate to the genome of *S. aureus* such as those set out elsewhere herein.)

It is clear that the etiology of diseases mediated or exacerbated by *S. aureus* infection involves the programmed expression of *S. aureus* genes, and that characterizing the genes and their patterns of expression would add dramatically to our understanding of the organism and its host interactions. Knowledge of *S. aureus* genes and genomic organization would dramatically improve understanding of disease etiology and lead to improved and new ways of preventing, ameliorating, arresting and reversing diseases. Moreover, characterized genes and genomic fragments of *S. aureus* would provide reagents for, among other things, detecting, characterizing and controlling *S. aureus* infections. There is a need therefore to characterize the genome of *S. aureus* and for polynucleotides and sequences of this organism.

The present invention is based on the sequencing of fragments of the *Staphylococcus aureus* genome. The primary nucleotide sequences which were generated are provided in SEQ ID NOS: 1-5,191.

The present invention provides the nucleotide sequence of several thousand contigs of the *Staphylococcus aureus* genome, which are listed in tables below and set out in the Sequence Listing submitted herewith, and representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan. In one embodiment, the present invention is provided as contiguous strings of primary sequence information corresponding to the nucleotide sequences depicted in SEQ ID NOS:1-5,191.

The present invention further provides nucleotide sequences which are at least 95%, preferably 99% and most preferably 99.9%, identical to the nucleotide sequences of SEQ ID NOS:1-5,191.

The nucleotide sequence of SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence which is at least 95%, preferably 99% and most preferably 99.9%, identical to the nucleotide sequence of SEQ ID NOS:1-5,191 may be provided in a variety of mediums to facilitate its use. In one application of this embodiment, the sequences of the present invention are recorded on computer readable media. Such media includes, but is not limited to:magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media.

The present invention further provides systems, particularly computer-based systems which contain the sequence

information herein described stored in a data storage means. Such systems are designed to identify commercially important fragments of the *Staphylococcus aureus* genome.

Another embodiment of the present invention is directed to fragments, preferably isolated fragments, of the *Staphylococcus aureus* genome having particular structural or functional attributes. Such fragments of the *Staphylococcus aureus* genome of the present invention include, but are not limited to, fragments which encode peptides, hereinafter referred to as open reading frames or ORFs," fragments which modulate the expression of an operably linked ORF, hereinafter referred to as expression modulating fragments or EMFs," and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in a sample, hereinafter referred to as diagnostic fragments or "DFs."

Each of the ORFs in fragments of the *Staphylococcus aureus* genome disclosed in Tables 1-3, and the EMFs found 5' to the ORFs, can be used in numerous ways as polynucleotide reagents. For instance, the sequences can be used as diagnostic probes or amplification primers for detecting or determining the presence of a specific microbe in a sample, to selectively control gene expression in a host and in the production of polypeptides, such as polypeptides encoded by ORFs of the present invention, particular those polypeptides that have a pharmacological activity.

The present invention further includes recombinant constructs comprising one or more fragments of the *Staphylococcus aureus* genome of the present invention. The recombinant constructs of the present invention comprise vectors, such as a plasmid or viral vector, into which a fragment of the *Staphylococcus aureus* has been inserted.

The present invention further provides host cells containing any of the isolated fragments of the *Staphylococcus aureus* genome of the present invention. The host cells can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic cell, such as a yeast cell, or a procaryotic cell such as a bacterial cell.

The present invention is further directed to polypeptides and proteins, preferably isolated polypeptides and proteins, encoded by ORFs of the present invention. A variety of methods, well known to those of skill in the art, routinely may be utilized to obtain any of the polypeptides and proteins of the present invention. For instance, polypeptides and proteins of the present invention having relatively short, simple amino acid sequences readily can be synthesized using commercially available automated peptide synthesizers. Polypeptides and proteins of the present invention also may be purified from bacterial cells which naturally produce the protein. Yet another alternative is to purify polypeptide and proteins of the present invention can from cells which have been altered to express them.

The invention further provides polypeptides, preferably isolated polypeptides, comprising *Staphylococcus aureus* epitopes and vaccine compositions comprising such polypeptides. Also provided are methods for vacciniating an individual against *Staphylococcus aureus* infection.

The invention further provides methods of obtaining homologs of the fragments of the *Staphylococcus aureus* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. Specifically, by using the nucleotide and amino acid sequences disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

The invention further provides antibodies which selectively bind polypeptides and proteins of the present invention. Such antibodies include both monoclonal and polyclonal antibodies.

The invention further provides hybridomas which produce the above-described antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

The present invention further provides methods of identifying test samples derived from cells which express one of the ORFs of the present invention, or a homolog thereof. Such methods comprise incubating a test sample with one or more of the antibodies of the present invention, or one or more of the Dfs or antigens of the present invention, under conditions which allow a skilled artisan to determine if the sample contains the ORF or product produced therefrom.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the above-described assays.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the antibodies, antigens, or one of the DFs of the present invention; and (b) one or more other containers comprising one or more of the following:wash reagents, reagents capable of detecting presence of bound antibodies, antigens or hybridized DFs.

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents capable of binding to a polypeptide or protein encoded by one of the ORFs of the present invention. Specifically, such agents include, as further described below, antibodies, peptides, carbohydrates, pharmaceutical agents and the like. Such methods comprise steps of: (a)contacting an agent with an isolated protein encoded by one of the ORFs of the present invention; and (b)determining whether the agent binds to said protein.

The present genomic sequences of *Staphylococcus aureus* will be of great value to all laboratories working with this organism and for a variety of commercial purposes. Many fragments of the *Staphylococcus aureus* genome will be immediately identified by similarity searches against GenBank or protein databases and will be of immediate value to *Staphylococcus aureus* researchers and for immediate commercial value for the production of proteins or to control gene expression.

The methodology and technology for elucidating extensive genomic sequences of bacterial and other genomes

has and will greatly enhance the ability to analyze and understand chromosomal organization. In particular, sequenced contigs and genomes will provide the models for developing tools for the analysis of chromosome structure and function, including the ability to identify genes within large segments of genomic DNA, the structure, position, and spacing of regulatory elements, the identification of genes with potential industrial applications, and the ability to do comparative genomic and molecular phylogeny.

FIGURE 1 is a block diagram of a computer system (102) that can be used to implement computer-based systems of present invention.

FIGURE 2 is a schematic diagram depicting the data flow and computer programs used to collect, assemble, edit and annotate the contigs of the *Staphylococcus aureus* genome of the present invention. Both Macintosh and Unix platforms are used to handle the AB 373 and 377 sequence data files, largely as described in Kerlavage *et al., Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences,* 585, IEEE Computer Society Press, Washington D.C. (1993). Factura (AB) is a Macintosh program designed for automatic vector sequence removal and end-trimming of sequence files. The program Loadis runs on a Macintosh platform and parses the feature data extracted from the sequence files by Factura to the Unix based *Staphylococcus aureus* relational database. Assembly of contigs (and whole genome sequences) is accomplished by retrieving a specific set of sequence files and their associated features using extrseq, a Unix utility for retrieving sequences from an SQL database. The resulting sequence file is processed by seq_filter to trim portions of the sequences with more than 2% ambiguous nucleotides. The sequence files were assembled using TIGR Assembler, an assembly engine designed at The Institute for Genomic Research (TIGR") for rapid and accurate assembly of thousands of sequence fragments. The collection of contigs generated by the assembly step is loaded into the database with the lassie program. Identification of open reading frames (ORFs) is accomplished by processing contigs with zorf. The ORFs are searched against *S. aureus* sequences from Genbank and against all protein sequences using the BLASTN and BLASTP programs, described in Altschul *et al., J. Mol. Biol.* 215: 403-410 (1990)). Results of the ORF determination and similarity searching steps were loaded into the database. As described below, some results of the determination and the searches are set out in Tables 1-3..

The present invention is based on the sequencing of fragments of the *Staphylococcus aureus* genome and analysis of the sequences. The primary nucleotide sequences generated by sequencing the fragments are provided in SEQ ID NOS:1-5,191. (As used herein, the "primary sequence" refers to the nucleotide sequence represented by the IUPAC nomenclature system.)

In addition to the aforementioned *Staphylococcus aureus* polynucleotide and polynucleotide sequences, the present invention provides the nucleotide sequences of SEQ ID NOS:1-5,191, or representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan.

As used herein, a "representative fragment of the nucleotide sequence depicted in SEQ ID NOS:1-5,191" refers to any portion of the SEQ ID NOS:1-5,191 which is not presently represented within a publicly available database. Preferred representative fragments of the present invention are *Staphylococcus aureus* open reading frames (ORFs"), expression modulating fragment (EMFs") and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in sample ("DFs"). A non-limiting identification of preferred representative fragments is provided in Tables 1-3.

As discussed in detail below, the information provided in SEQ ID NOS:1-5,191 and in Tables 1-3 together with routine cloning, synthesis, sequencing and assay methods will enable those skilled in the art to clone and sequence all "representative fragments" of interest, including open reading frames encoding a large variety of *Staphylococcus aureus* proteins.

While the presently disclosed sequences of SEQ ID NOS:1-5,191 are highly accurate, sequencing techniques are not perfect and, in relatively rare instances, further investigation of a fragment or sequence of the invention may reveal a nucleotide sequence error present in a nucleotide sequence disclosed in SEQ ID NOS:1-5,191. However, once the present invention is made available (*i.e.,* once the information in SEQ ID NOS:1-5,191 and Tables 1-3 has been made available), resolving a rare sequencing error in SEQ ID NOS:1-5,191 will be well within the skill of the art. The present disclosure makes available sufficient sequence information to allow any of the described contigs or portions thereof to be obtained readily by straightforward application of routine techniques. Further sequencing of such polynucleotide may proceed in like manner using manual and automated sequencing methods which are employed ubiquitous in the art. Nucleotide sequence editing software is publicly available. For example, Applied Biosystem's (AB) AutoAssembler can be used as an aid during visual inspection of nucleotide sequences. By employing such routine techniques potential errors readily may be identified and the correct sequence then may be ascertained by targeting further sequencing effort, also of a routine nature, to the region containing the potential error.

Even if all of the very rare sequencing errors in SEQ ID NOS:1-5,191 were corrected, the resulting nucleotide sequences would still be at least 95% identical, nearly all would be at least 99% identical, and the great majority would be at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191.

As discussed elsewhere hererin, polynucleotides of the present invention readily may be obtained by routine application of well known and standard procedures for cloning and sequencing DNA. Detailed methods for obtaining

libraries and for sequencing are provided below, for instance. A wide variety of *Staphylococcus aureus* strains that can be used to prepare *S aureus* genomic DNA for cloning and for obtaining polynucleotides of the present invention are available to the public from recognized depository institutions, such as the American Type Culture Collection (ATCC").

The nucleotide sequences of the genomes from different strains of *Staphylococcus aureus* differ somewhat. However, the nucleotide sequences of the genomes of all *Staphylococcus aureus* strains will be at least 95% identical, in corresponding part, to the nucleotide sequences provided in SEQ ID NOS:1-5,191. Nearly all will be at least 99% identical and the great majority will be 99.9% identical.

Thus, the present invention further provides nucleotide sequences which are at least 95%, preferably 99% and most preferably 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191, in a form which can be readily used, analyzed and interpreted by the skilled artisan.

Methods for determining whether a nucleotide sequence is at least 95%, at least 99% or at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191 are routine and readily available to the skilled artisan. For example, the well known fasta algorithm described in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85: 2444 (1988) can be used to generate the percent identity of nucleotide sequences. The BLASTN program also can be used to generate an identity score of polynucleotides compared to one another.

COMPUTER RELATED EMBODIMENTS

The nucleotide sequences provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide sequence of SEQ ID NOS:1-5,191 may be "provided" in a variety of mediums to facilitate use thereof. As used herein, Òprovided" refers to a manufacture, other than an isolated nucleic acid molecule, which contains a nucleotide sequence of the present invention; *i.e.,* a nucleotide sequence provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide of SEQ ID NOS:1-5,191. Such a manufacture provides a large portion of the *Staphylococcus aureus* genome and parts thereof (*e.g.,* a *Staphylococcus aureus* open reading frame (ORF)) in a form which allows a skilled artisan to examine the manufacture using means not directly applicable to examining the *Staphylococcus aureus* genome or a subset thereof as it exists in nature or in purified form.

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD- ROM; electrical storage media such as RAM and ROM; and hybrids of these categories, such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. Likewise, it will be clear to those of skill how additional computer readable media that may be developed also can be used to create analogous manufactures having recorded thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently know methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially- available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data-processor structuring formats (*e.g.,* text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Thus, by providing in computer readable form the nucleotide sequences of SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a sequence of SEQ ID NOS:1-5,191 the present invention enables the skilled artisan routinely to access the provided sequence information for a wide variety of purposes.

The examples which follow demonstrate how software which implements the BLAST (Altschul *et al.,* J. Mol. Biol. 215:403410 (1990)) and BLAZE (Brutlag *et al.,* Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system was used to identify open reading frames (ORFs) within the *Staphylococcus aureus* genome which contain homology to ORFs or proteins from both *Staphylococcus aureus* and from other organisms. Among the ORFs discussed

herein are protein encoding fragments of the *Staphylococcus aureus* genome useful in producing commercially important proteins, such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

The present invention further provides systems, particularly computer-based systems, which contain the sequence information described herein. Such systems are designed to identify, among other things, commercially important fragments of the *Staphylococcus aureus* genome.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention.

As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means.

As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer- based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the present genomic sequences which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the *Staphylococcus aureus* genomic sequences possessing varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the *Staphylococcus aureus* genome. In the present examples, implementing software which implement the BLAST and BLAZE algorithms, described in Altschul *et al., J. Mol. Biol.* 215: 403-410 (1990), was used to identify open reading frames within the *Staphylococcus aureus* genome. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer-based systems of the present invention. Of course, suitable proprietary systems that may be known to those of skill also may be employed in this regard.

Figure 1 provides a block diagram of a computer system illustrative of embodiments of this aspect of present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM) and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, *etc.* A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, *etc.*) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the data from the removable medium storage device 114, once it is inserted into the removable medium storage device 114.

A nucleotide sequence of the present invention may be stored in a well known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. During execution, software for

accessing and processing the genomic sequence (such as search tools, comparing tools, *etc.*) reside in main memory 108, in accordance with the requirements and operating parameters of the operating system, the hardware system and the software program or programs.

BIOCHEMICAL EMBODIMENTS

Other embodiments of the present invention are directed to fragments of the *Staphylococcus aureus* genome, preferably to isolated fragments. The fragments of the *Staphylococcus aureus* genome of the present invention include, but are not limited to fragments which encode peptides, hereinafter open reading frames (ORFs), fragments which modulate the expression of an operably linked ORF, hereinafter expression modulating fragments (EMFs) and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in a sample, hereinafter diagnostic fragments (DFs).

As used herein, an "isolated nucleic acid molecule" or an "isolated fragment of the *Staphylococcus aureus* genome" refers to a nucleic acid molecule possessing a specific nucleotide sequence which has been subjected to purification means to reduce, from the composition, the number of compounds which are normally associated with the composition. Particularly, the term refers to the nucleic acid molecules having the sequences set out in SEQ ID NOS:1-5,191, to representative fragments thereof as described above, to polynucleotides at least 95%, preferably at least 99% and especially preferably at least 99.9% identical in sequence thereto, also as set out above.

A variety of purification means can be used to generated the isolated fragments of the present invention. These include, but are not limited to methods which separate constituents of a solution based on charge, solubility, or size.

In one embodiment, *Staphylococcus aureus* DNA can be mechanically sheared to produce fragments of 15-20 kb in length. These fragments can then be used to generate an *Staphylococcus aureus* library by inserting them into lambda clones as described in the Examples below. Primers flanking, for example, an ORF, such as those enumerated in Tables 1-3 can then be generated using nucleotide sequence information provided in SEQ ID NOS: 1-5,191. Well known and routine techniques of PCR cloning then can be used to isolate the ORF from the lambda DNA library of Staphylococcus aureus genomic DNA. Thus, given the availability of SEQ ID NOS:1-5,191, the information in Tables 1, 2 and 3, and the information that may be obtained readily by analysis of the sequences of SEQ ID NOS:1-5,191 using methods set out above, those of skill will be enabled by the present disclosure to isolate any ORF-containing or other nucleic acid fragment of the present invention.

The isolated nucleic acid molecules of the present invention include, but are not limited to single stranded and double stranded DNA, and single stranded RNA.

As used herein, an "open reading frame," ORF, means a series of triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

Tables 1, 2 and 3 list ORFs in the *Staphylococcus aureus* genomic contigs of the present invention that were identified as putative coding regions by the GeneMark software using organism-specific second-order Markov probability transition matrices. It will be appreciated that other criteria can be used, in accordance with well known analytical methods, such as those discussed herein, to generate more inclusive, more restrictive or more selective lists.

Table 1 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that are at least 80 amino acids long and over a continuous region of at least 50 bases which are 95% or more identical (by BLAST analysis) to an *S. aureus* nucleotide sequence available through Genbank in November 1996.

Table 2 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that are not in Table 1 and match, with a BLASTP probability score of 0.01 or less, a polypeptide sequence available through Genbank by September 1996.

Table 3 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that do not match significantly, by BLASTP analysis, a polypeptide sequence available through Genbank by September 1996.

In each table, the first and second columns identify the ORF by, respectively, contig number and ORF number within the contig; the third column indicates the reading frame, taking the first 5' nucleotide of the contig as the start of the +1 frame; the fourth column indicates the first nucleotide of the ORF, counting from the 5' end of the contig strand; and the fifth column indicates the length of each ORF in nucleotides.

In Tables 1 and 2, column six, lists the Reference" for the closest matching sequence available through Genbank. These reference numbers are the databases entry numbers commonly used by those of skill in the art, who will be familiar with their denominators. Descriptions of the numenclature are available from the National Center for Biotechnology Information. Column seven in Tables 1 and 2 provides the gene name" of the matching sequence; column eight provides the BLAST identity" score from the comparison of the ORF and the homologous gene; and column nine indicates the length in nucleotides of the highest scoring segment pair" identified by the BLAST identity analysis.

In Table 3, the last column, column six, indicates the length of each ORF in amino acid residues.

The concepts of percent identity and percent similarity of two polypeptide sequences is well understood in the art. For example, two polypeptides 10 amino acids in length which differ at three amino acid positions (*e.g.*, at positions

1, 3 and 5) are said to have a percent identity of 70%. However, the same two polypeptides would be deemed to have a percent similarity of 80% if, for example at position 5, the amino acids moieties, although not identical, were "similar" (*i.e.,* possessed similar biochemical characteristics). Many programs for analysis of nucleotide or amino acid sequence similarity, such as fasta and BLAST specifically list per cent identity of a matching region as an output parameter. Thus, for instance, Tables 1 and 2 herein enumerate the per cent identity" of the highest scoring segment pair" in each ORF and its listed relative. Further details concerning the algorithms and criteria used for homology searches are provided below and are described in the pertinent literature highlighted by the citations provided below.

It will be appreciated that other criteria can be used to generate more inclusive and more exclusive listings of the types set out in the tables. As those of skill will appreciate, narrow and broad searches both are useful. Thus, a skilled artisan can readily identify ORFs in contigs of the *Staphylococcus aureus* genome other than those listed in Tables 1-3, such as ORFs which are overlapping or encoded by the opposite strand of an identified ORF in addition to those ascertainable using the computer-based systems of the present invention.

As used herein, an "expression modulating fragment," EMF, means a series of nucleotide molecules which modulates the expression of an operably linked ORF or EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are fragments which induce the expression or an operably linked ORF in response to a specific regulatory factor or physiological event.

EMF sequences can be identified within the contigs of the *Staphylococcus aureus* genome by their proximity to the ORFs provided in Tables 1-3. An intergenic segment, or a fragment of the intergenic segment, from about 10 to 200 nucleotides in length, taken from any one of the ORFs of Tables 1-3 will modulate the expression of an operably linked ORF in a fashion similar to that found with the naturally linked ORF sequence. As used herein, an "intergenic segment" refers to fragments of the *Staphylococcus aureus* genome which are between two ORF(s) herein described. EMFs also can be identified using known EMFs as a target sequence or target motif in the computer-based systems of the present invention. Further, the two methods can be combined and used together.

The presence and activity of an EMF can be confirmed using an EMF trap vector. An EMF trap vector contains a cloning site linked to a marker sequence. A marker sequence encodes an identifiable phenotype, such as antibiotic resistance or a complementing nutrition auxotrophic factor, which can be identified or assayed when the EMF trap vector is placed within an appropriate host under appropriate conditions. As described above, a EMF will modulate the expression of an operably linked marker sequence. A more detailed discussion of various marker sequences is provided below.

A sequence which is suspected as being an EMF is cloned in all three reading frames in one or more restriction sites upstream from the marker sequence in the EMF trap vector. The vector is then transformed into an appropriate host using known procedures and the phenotype of the transformed host in examined under appropriate conditions. As described above, an EMF will modulate the expression of an operably linked marker sequence.

As used herein, a "diagnostic fragment," DF, means a series of nucleotide molecules which selectively hybridize to *Staphylococcus aureus* sequences. DFs can be readily identified by identifying unique sequences within contigs of the *Staphylococcus aureus* genome, such as by using well-known computer analysis software, and by generating and testing probes or amplification primers consisting of the DF sequence in an appropriate diagnostic format which determines amplification or hybridization selectivity.

The sequences falling within the scope of the present invention are not limited to the specific sequences herein described, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequences provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably 99% and most preferably 99.9% identical to SEQ ID NOS:1-5,191, with a sequence from another isolate of the same species.

Furthermore, to accomodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the nucleic acid sequences mentioned above. In other words, in the coding region of an ORF, substitution of one codon for another which encodes the same amino acid is expressly contemplated.

Any specific sequence disclosed herein can be readily screened for errors by resequencing a particular fragment, such as an ORF, in both directions (*i.e.,* sequence both strands). Alternatively, error screening can be performed by sequencing corresponding polynucleotides of *Staphylococcus aureus* origin isolated by using part or all of the fragments in question as a probe or primer.

Each of the ORFs of the *Staphylococcus aureus* genome disclosed in Tables 1, 2 and 3, and the EMFs found 5' to the ORFs, can be used as polynucleotide reagents in numerous ways. For example, the sequences can be used as diagnostic probes or diagnostic amplification primers to detect the presence of a specific microbe in a sample, particular *Staphylococcus aureus.* Especially preferred in this regard are ORF such as those of Table 3, which do not match previously characterized sequences from other organisms and thus are most likely to be highly selective for *Staphylococcus aureus.* Also particularly preferred are ORFs that can be used to distinguish between strains of *Sta-*

phylococcus aureus, particularly those that distinguish medically important strain, such as drug-resistant strains.

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Triple helix- formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Information from the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription, for triple-helix formation, or to the mRNA itself, for antisense inhibition. Both techniques have been demonstrated to be effective in model systems, and the requisite techniques are well known and involve routine procedures. Triple helix techniques are discussed in, for example, Lee *et al., Nucl. Acids Res.* 6: 3073 (1979); Cooney *et al., Science* 241: 456 (1988); and Dervan *et al., Science* 251: 1360 (1991). Antisense techniques in general are discussed in, for instance, Okano, *J. Neurochem.* 56: 560 (1991) and OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION, CRC Press, Boca Raton, FL (1988)).

The present invention further provides recombinant constructs comprising one or more fragments of the *Staphylococcus aureus* genomic fragments and contigs of the present invention. Certain preferred recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a fragment of the *Staphylococcus aureus* genome has been inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. For vectors comprising the EMFs of the present invention, the vector may further comprise a marker sequence or heterologous ORF operably linked to the EMF.

Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Useful bacterial vectors include phagescript, PsiX174, pBluescript SK and KS (+ and -), pNH8a, pNH16a, pNH18a, pNH46a (available from Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (available from Pharmacia). Useful eukaryotic vectors include pWLneo, pSV2cat, pOG44, pXT1, pSG (available from Stratagene) pSVK3, pBPV, pMSG, pSVL (available from Pharmacia).

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein- I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

The present invention further provides host cells containing any one of the isolated fragments of the *Staphylococcus aureus* genomic fragments and contigs of the present invention, wherein the fragment has been introduced into the host cell using known methods. The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or a procaryotic cell, such as a bacterial cell.

A polynucleotide of the present invention, such as a recombinant construct comprising an ORF of the present invention, may be introduced into the host by a variety of well established techniques that are standard in the art, such as calcium phosphate transfection, DEAE, dextran mediated transfection and electroporation, which are described in, for instance, Davis, L. *et al.,* BASIC METHODS IN MOLECULAR BIOLOGY (1986).

A host cell containing one of the fragments of the *Staphylococcus aureus* genomic fragments and contigs of the present invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (*e.g.,* an ORF) by nucleotide sequence but, due to the degeneracy of the Genetic Code, encode an identical polypeptide sequence.

Preferred nucleic acid fragments of the present invention are the ORFs depicted in Tables 2 and 3 which encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. Such short fragments as may be obtained most readily by synthesis are useful, for example, in generating antibodies against the native polypeptide, as discussed further below.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily employ well-known methods for isolating polpeptides and proteins to isolate and purify polypeptides or proteins of the present invention produced naturally by a bacterial strain, or by other methods. Methods for isolation and purification that can be employed in this regard include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immu-

no-affinity chromatography.

The polypeptides and proteins of the present invention also can be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. Those skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, CV-1 cell, COS cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level.

"Recombinant," as used herein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (*e.g.*, yeast) expression systems. As a product, "recombinant microbial"defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g., E. coli,* will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells.

"Nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides. Generally, DNA segments encoding the polypeptides and proteins provided by this invention are assembled from fragments of the *Staphylococcus aureus* genome and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic gene which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon.

"Recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. The expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic regulatory elements necessary for gene expression in the host, including elements required to initiate and maintain transcription at a level sufficient for suitable expression of the desired polypeptide, including, for example, promoters and, where necessary, an enhancers and a polyadenylation signal; (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate signals to initiate translation at the beginning of the desired coding region and terminate translation at its end. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

"Recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extra chromosomally. The cells can be prokaryotic or eukaryotic. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described in Sambrook *et al.,* MOLECULAR CLONING:A LABORATORY MANUAL, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference in its entirety.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.,* the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), alpha-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and, when desirable, provide amplification within the host.

Suitable prokaryotic hosts for transformation include strains of *Staphylococcus aureus, E. coli, B. subtilis, Salmonella typhimurium* and various species within the genera *Pseudomonas, Streptomyces,* and *Staphylococcus.* Others

may, also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (available form Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (available from Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter, where it is inducible, is derepressed or induced by appropriate means (*e.g.,* temperature shift or chemical induction) and cells are cultured for an additional period to provide for expression of the induced gene product. Thereafter cells are typically harvested, generally by centrifugation, disrupted to release expressed protein, generally by physical or chemical means, and the resulting crude extract is retained for further purification.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described in Gluzman, *Cell* 23: 175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines.

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Recombinant polypeptides and proteins produced in bacterial culture is usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

An additional aspect of the invention includes Staphylococcus aureus polypeptides which are useful as immunodiagnostic antigens and/or immunoprotective vaccines, collectively "immunologically useful polypeptides". Such immunologically useful polypeptides may be selected from the ORFs disclosed herein based on techniques well known in the art and described elsewhere herein. The inventors have used the following criteria to select several immunologically useful polypeptides:

As is known in the art, an amino terminal type I signal sequence directs a nascent protein across the plasma and outer membranes to the exterior of the bacterial cell. Such outermembrane polypeptides are expected to be immunologically useful. According to Izard, J. W. et al., Mol. Microbiol. 13, 765-773; (1994), polypeptides containing type I signal sequences contain the following physical attributes: The length of the type I signal sequence is approximately 15 to 25 primarily hydrophobic amino acid residues with a net positive charge in the extreme amino terminus; the central region of the signal sequence must adopt an alpha-helical conformation in a hydrophobic environment; and the region surrounding the actual site of cleavage is ideally six residues long, with small side-chain amino acids in the -1 and -3 positions.

Also known in the art is the type IV signal sequence which is an example of the several types of functional signal sequences which exist in addition to the type I signal sequence detailed above. Although functionally related, the type IV signal sequence possesses a unique set of biochemical and physical attributes (Strom, M. S. and Lory, S., J. Bacteriol. 174, 7345-7351; 1992)). These are typically six to eight amino acids with a net basic charge followed by an additional sixteen to thirty primarily hydrophobic residues. The cleavage site of a type IV signal sequence is typically after the initial six to eight amino acids at the extreme amino terminus. In addition, all type IV signal sequences contain a phenylalanine residue at the +1 site relative to the cleavage site.

Studies of the cleavage sites of twenty-six bacterial lipoprotein precursors has allowed the definition of a consensus amino acid sequence for lipoprotein cleavage. Nearly three-fourths of the bacterial lipoprotein precursors examined contained the sequence L-(A,S)-(G,A)-C at positions -3 to +1, relative to the point of cleavage (Hayashi, S. and Wu, H. C. Lipoproteins in bacteria. J Bioenerg. Biomembr. 22, 451-471; 1990).

It well known that most anchored proteins found on the surface of gram-positive bacteria possess a highly conserved carboxy terminal sequence. More than fifty such proteins from organisms such as *S. pyogenes, S. mutans, E. faecalis, S. pneumoniae,* and others, have been identified based on their extracellular location and carboxy terminal amino acid sequence (Fischetti, V. A. Gram-positive commensal bacteria deliver antigens to elicit mucosal and systemic immunity. ASM News 62, 405410; 1996). The conserved region is comprised of six charged amino acids at the extreme carboxy terminus coupled to 15-20 hydrophobic amino acids presumed to function as a transmembrane domain. Immediately adjacent to the transmembrane domain is a six amino acid sequence conserved in nearly all proteins ex-

amined. The amino acid sequence of this region is L-P-X-T-G-X, where X is any amino acid.

Amino acid sequence similarities to proteins of known function by BLAST enables the assignment of putative functions to novel amino acid sequences and allows for the selection of proteins thought to function outside the cell wall. Such proteins are well known in the art and include "lipoprotein", "periplasmic", or "antigen".

An algorithm for selecting antigenic and immunogenic *Staphylococcus aureus* polypeptides including the foregoing criteria was developed by the present inventors. Use of the algorithm by the inventors to select immunologically useful *Staphylococcus aureus* polypeptides resulted in the selection of several ORFs which are predicted to be outermembrane-associated proteins. These proteins are identified in Table 4, below, and shown in the Sequence Listing as SEQ ID NOS:5,192 to 5,255. Thus the amino acid sequence of each of several antigenic *Staphylococcus aureus* polypeptides listed in Table 4 can be determined, for example, by locating the amino acid sequence of the ORF in the Sequence Listing. Likewise the polynucleotide sequence encoding each ORF can be found by locating the corresponding polynucleotide SEQ ID in Tables 1, 2, or 3, and finding the corresponding nucleotide sequence in the sequence listing.

As will be appreciated by those of ordinary skill in the art, although a polypeptide representing an entire ORF may be the closest approximation to a protein found *in vivo*, it is not always technically practical to express a complete ORF *in vitro*. It may be very challenging to express and purify a highly hydrophobic protein by common laboratory methods. As a result, the immunologically useful polypeptides described herein as SEQ ID NOS:5,192-5,255 may have been modified slightly to simplify the production of recombinant protein, and are the preferred embodiments. In general, nucleotide sequences which encode highly hydrophobic domains, such as those found at the amino terminal signal sequence, are excluded for enhanced *in vitro* expression of the polypeptides. Furthermore, any highly hydrophobic amino acid sequences occurring at the carboxy terminus are also excluded. Such truncated polypeptides include for example the mature forms of the polypeptides expected to exist in nature.

Those of ordinary skill in the art can identify soluble portions the polypeptide identified in Table 4, and in the case of truncated polypeptides sequences shown as SEQ ID NOS:5,192-5,255, may obtain the complete predicted amino acid sequence of each polypeptide by translating the corresponding polynucleotides sequences of the corresponding ORF listed in Tables 1,2 and 3 and found in the sequence listing.

Accordingly, polypeptides comprising the complete amino acid of an immunologically useful polypeptide selected from the group of polypeptides encoded by the ORFs identified in Table 4, or an amino acid sequence at least 95% identical thereto, preferably at least 97% identical thereto, and most preferably at least 99% identical thereto form an embodiment of the invention; in addition polypeptides comprising an amino acid sequence selected from the group of amino acid sequences shown in the sequence listing as SEQ ID NOS:5,191-5,255, or an amino acid sequence at least 95% identical thereto, preferably at least 97% identical thereto and most preferably at least 99% identical thereto, form an embodiment of the invention. Polynucleotides encoding the foregoing polypeptides also form part of the present invention.

In another aspect, the invention provides a peptide or polypeptide comprising an epitope-bearing portion of a polypeptide of the invention, particularly those epitope-bearing portions (antigenic regions) identified in Table 4. The epitope-bearing portion is an immunogenic or antigenic epitope of a polypeptide of the invention. An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response when the whole protein is the immunogen. On the other hand, a region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." The number of immunogenic epitopes of a protein generally is less than the number of antigenic epitopes. See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998- 4002 (1983).

As to the selection of peptides or polypeptides bearing an antigenic epitope (i.e., that contain a region of a protein molecule to which an antibody can bind), it is well known in that art that relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein. See, for instance, Sutcliffe, J. G., Shinnick, T. M., Green, N. and Learner, R. A. (1983) "Antibodies that react with predetermined sites on proteins", Science, 219:660-666. Peptides capable of eliciting protein-reactive sera are frequently represented in the primary sequence of a protein, can be characterized by a set of simple chemical rules, and are confined neither to immunodominant regions of intact proteins (i.e., immunogenic epitopes) nor to the amino or carboxyl terminals. Antigenic epitope-bearing peptides and polypeptides of the invention are therefore useful to raise antibodies, including monoclonal antibodies, that bind specifically to a polypeptide of the invention. See, for instance, Wilson et al., Cell 37:767-778 (1984) at 777.

Antigenic epitope-bearing peptides and polypeptides of the invention preferably contain a sequence of at least seven, more preferably at least nine and most preferably between about 15 to about 30 amino acids contained within the amino acid sequence of a polypeptide of the invention. Non-limiting examples of antigenic polypeptides or peptides that can be used to generate S. aureus specific antibodies include: a polypeptide comprising peptides shown in Table 4 below. These polypeptide fragments have been determined to bear antigenic epitopes of indicated S. aureus proteins by the analysis of the Jameson-Wolf antigenic index, a representative sample of which is shown in Figure 3.

The epitope-bearing peptides and polypeptides of the invention may be produced by any conventional means. See, e.g., Houghten, R. A. (1985) General method for the rapid solid-phase synthesis of large numbers of peptides:

specificity of antigen-antibody interaction at the level of individual amino acids. Proc. Natl. Acad. Sci. USA 82: 5131-5135; this "Simultaneous Multiple Peptide Synthesis (SMPS)" process is further described in U.S. Patent No. 4,631,211 to Houghten et al. (1986). Epitope-bearing peptides and polypeptides of the invention are used to induce antibodies according to methods well known in the art. See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).

Immunogenic epitope-bearing peptides of the invention, i.e., those parts of a protein that elicit an antibody response when the whole protein is the immunogen, are identified according to methods known in the art. See, for instance, Geysen et al., supra. Further still, U.S. Patent No. 5,194,392 to Geysen (1990) describes a general method of detecting or determining the sequence of monomers (amino acids or other compounds) which is a topological equivalent of the epitope (i.e., a "mimotope") which is complementary to a particular paratope (antigen binding site) of an antibody of interest. More generally, U.S. Patent No. 4,433,092 to Geysen (1989) describes a method of detecting or determining a sequence of monomers which is a topographical equivalent of a ligand which is complementary to the ligand binding site of a particular receptor of interest. Similarly, U.S. Patent No. 5,480,971 to Houghten, R. A. et al. (1996) on Peralkylated Oligopeptide Mixtures discloses linear C1-C7-alkyl peralkylated oligopeptides and sets and libraries of such peptides, as well as methods for using such oligopeptide sets and libraries for determining the sequence of a peralkylated oligopeptide that preferentially binds to an acceptor molecule of interest. Thus, non-peptide analogs of the epitope-bearing peptides of the invention also can be made routinely by these methods.

Table 4 lists immunologically useful polypeptides identified by an algorithm which locates novel *Staphylococcus aureus* outermembrane proteins, as is described above. Also listed are epitopes or "antigenic regions" of each of the identified polypeptides. The antigenic regions, or epitopes, are delineated by two numbers x-y, where x is the number of the first amino acid in the open reading frame included within the epitope and y is the number of the last amino acid in the open reading frame included within the epitope. For example, the first epitope in ORF 168-6 is comprised of amino acids 36 to 45 of SEQ ID NO:5,192, as is described in Table 4. The inventors have identified several epitopes for each of the antigenic polypeptides identified in Table 4. Accordingly, forming part of the present invention are polypeptides comprising an amino acid sequence of one or more antigenic regions identified in Table 4. The invention further provides polynucleotides encoding such polypeptides.

The present invention further includes isolated polypeptides, proteins and nucleic acid molecules which are substantially equivalent to those herein described. As used herein, substantially equivalent can refer both to nucleic acid and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between reference and subject sequences. For purposes of the present invention, sequences having equivalent biological activity, and equivalent expression characteristics are considered substantially equivalent. For purposes of determining equivalence, truncation of the mature sequence should be disregarded.

The invention further provides methods of obtaining homologs from other strains of *Staphylococcus aureus,* of the fragments of the *Staphylococcus aureus* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. As used herein, a sequence or protein of *Staphylococcus aureus* is defined as a homolog of a fragment of the *Staphylococcus aureus* fragments or contigs or a protein encoded by one of the ORFs of the present invention, if it shares significant homology to one of the fragments of the *Staphylococcus aureus* genome of the present invention or a protein encoded by one of the ORFs of the present invention. Specifically, by using the sequence disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

As used herein, two nucleic acid molecules or proteins are said to "share significant homology" if the two contain regions which prossess greater than 85% sequence (amino acid or nucleic acid) homology. Preferred homologs in this regard are those with more than 90% homology. Especially preferred are those with 93% or more homology. Among especially preferred homologs those with 95% or more homology are particularly preferred. Very particularly preferred among these are those with 97% and even more particularly preferred among those are homologs with 99% or more homology. The most preferred homologs among these are those with 99.9% homology or more. It will be understood that, among measures of homology, identity is particularly preferred in this regard.

Region specific primers or probes derived from the nucleotide sequence provided in SEQ ID NOS:1-5,191 or from a nucleotide sequence at least 95%, particularly at least 99%, especially at least 99.5% identical to a sequence of SEQ ID NOS:1-5,191 can be used to prime DNA synthesis and PCR amplification, as well as to identify colonies containing cloned DNA encoding a homolog. Methods suitable to this aspect of the present invention are well known and have been described in great detail in many publications such as, for example, Innis *et al.,* PCR PROTOCOLS, Academic Press, San Diego, CA (1990)).

When using primers derived from SEQ ID NOS:1-5,191 or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-5,191, one skilled in the art will recognize that by employing high stringency conditions (*e.g.,* annealing at 50-60°C in 6X SSPC and 50% formamide, and washing at 50- 65°C in 0.5X SSPC) only sequences which are greater than 75% homologous to the primer will be amplified. By employing lower stringency

conditions (*e.g.,* hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences which are greater than 40-50% homologous to the primer will also be amplified.

When using DNA probes derived from SEQ ID NOS:1-5,191, or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-5,191, for colony/plaque hybridization, one skilled in the art will recognize that by employing high stringency conditions *(e.g.,* hybridizing at 50- 65°C in 5X SSPC and 50% formamide, and washing at 50- 65°C in 0.5X SSPC), sequences having regions which are greater than 90% homologous to the probe can be obtained, and that by employing lower stringency conditions (*e.g.,* hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences having regions which are greater than 35-45% homologous to the probe will be obtained.

Any organism can be used as the source for homologs of the present invention so long as the organism naturally expresses such a protein or contains genes encoding the same. The most preferred organism for isolating homologs are bacterias which are closely related to *Staphylococcus aureus.*

ILLUSTRATIVE USES OF COMPOSITIONS OF THE INVENTION

Each ORF provided in Tables 1 and 2 is identified with a function by homology to a known gene or polypeptide. As a result, one skilled in the art can use the polypeptides of the present invention for commercial, therapeutic and industrial purposes consistent with the type of putative identification of the polypeptide. Such identifications permit one skilled in the art to use the *Staphylococcus aureus* ORFs in a manner similar to the known type of sequences for which the identification is made; for example, to ferment a particular sugar source or to produce a particular metabolite. A variety of reviews illustrative of this aspect of the invention are available, including the following reviews on the industrial use of enzymes, for example, BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY HANDBOOK, 2nd Ed., Macmillan Publications, Ltd. NY (1991) and BIOCATALYSTS IN ORGANIC SYNTHESES, Tramper *et al.,* Eds., Elsevier Science Publishers, Amsterdam, The Netherlands (1985). A variety of exemplary uses that illustrate this and similar aspects of the present invention are discussed below.

1. Biosynthetic Enzymes

Open reading frames encoding proteins involved in mediating the catalytic reactions involved in intermediary and macromolecular metabolism, the biosynthesis of small molecules, cellular processes and other functions includes enzymes involved in the degradation of the intermediary products of metabolism, enzymes involved in central intermediary metabolism, enzymes involved in respiration, both aerobic and anaerobic, enzymes involved in fermentation, enzymes involved in ATP proton motor force conversion, enzymes involved in broad regulatory function, enzymes involved in amino acid synthesis, enzymes involved in nucleotide synthesis, enzymes involved in cofactor and vitamin synthesis, can be used for industrial biosynthesis.

The various metabolic pathways present in *Staphylococcus aureus* can be identified based on absolute nutritional requirements as well as by examining the various enzymes identified in Table 1-3 and SEQ ID NOS:1-5,191.

Of particular interest are polypeptides involved in the degradation of intermediary metabolites as well as non-macromolecular metabolism. Such enzymes include amylases, glucose oxidases, and catalase.

Proteolytic enzymes are another class of commercially important enzymes. Proteolytic enzymes find use in a number of industrial processes including the processing of flax and other vegetable fibers, in the extraction, clarification and depectinization of fruit juices, in the extraction of vegetables' oil and in the maceration of fruits and vegetables to give unicellular fruits. A detailed review of the proteolytic enzymes used in the food industry is provided in Rombouts *et al., Symbiosis* 21: 79 (1986) and Voragen *et al.* in BIOCATALYSTS IN AGRICULTURAL BIOTECHNOLOGY, Whitaker *et al.,* Eds., *American Chemical Society Symposium Series* 389: 93 (1989) .

The metabolism of sugars is an important aspect of the primary metabolism of *Staphylococcus aureus.* Enzymes involved in the degradation of sugars, such as, particularly, glucose, galactose, fructose and xylose, can be used in industrial fermentation. Some of the important sugar transforming enzymes, from a commercial viewpoint, include sugar isomerases such as glucose isomerase. Other metabolic enzymes have found commercial use such as glucose oxidases which produces ketogulonic acid (KGA). KGA is an intermediate in the commercial production of ascorbic acid using the Reichstein's procedure, as described in Krueger *et al., Biotechnology* 6(A), Rhine *et al.,* Eds., Verlag Press, Weinheim, Germany (1984).

Glucose oxidase (GOD) is commercially available and has been used in purified form as well as in an immobilized form for the deoxygenation of beer. See, for instance, Hartmeir *et al., Biotechnology Letters* 1: 21 (1979). The most important application of GOD is the industrial scale fermentation of gluconic acid. Market for gluconic acids which are used in the detergent, textile, leather, photographic, pharmaceutical, food, feed and concrete industry, as described, for example, in Bigelis *et al.,* beginning on page 357 in GENE MANIPULATIONS AND FUNGI; Benett *et al.,* Eds., Academic Press, New York (1985). In addition to industrial applications, GOD has found applications in medicine for

quantitative determination of glucose in body fluids recently in biotechnology for analyzing syrups from starch and cellulose hydrosylates. This application is described in Owusu *et al., Biochem. et Biophysica. Acta.* 872: 83 (1986), for instance.

The main sweetener used in the world today is sugar which comes from sugar beets and sugar cane. In the field of industrial enzymes, the glucose isomerase process shows the largest expansion in the market today. Initially, soluble enzymes were used and later immobilized enzymes were developed (Krueger *et al.,* Biotechnology, The Textbook of Industrial Microbiology, Sinauer Associated Incorporated, Sunderland, Massachusetts (1990)). Today, the use of glucose- produced high fructose syrups is by far the largest industrial business using immobilized enzymes. A review of the industrial use of these enzymes is provided by Jorgensen, Starch 40:307 (1988).

Proteinases, such as alkaline serine proteinases, are used as detergent additives and thus represent one of the largest volumes of microbial enzymes used in the industrial sector. Because of their industrial importance, there is a large body of published and unpublished information regarding the use of these enzymes in industrial processes. (See Faultman *et al.,* Acid Proteases Structure Function and Biology, Tang, J., ed., Plenum Press, New York (1977) and Godfrey *et al.,* Industrial Enzymes, MacMillan Publishers, Surrey, UK (1983) and Hepner *et al.,* Report Industrial Enzymes by 1990, Hel Hepner & Associates, London (1986)).

Another class of commercially usable proteins of the present invention are the microbial lipases, described by, for instance, Macrae *et al.,* Philosophical Transactions of the Chiral Society of London 310:227 (1985) and Poserke, Journal of the American Oil Chemist Society 61:1758 (1984). A major use of lipases is in the fat and oil industry for the production of neutral glycerides using lipase catalyzed inter-esterification of readily available triglycerides. Application of lipases include the use as a detergent additive to facilitate the removal of fats from fabrics in the course of the washing procedures.

The use of enzymes, and in particular microbial enzymes, as catalyst for key steps in the synthesis of complex organic molecules is gaining popularity at a great rate. One area of great interest is the preparation of chiral intermediates. Preparation of chiral intermediates is of interest to a wide range of synthetic chemists particularly those scientists involved with the preparation of new pharmaceuticals, agrochemicals, fragrances and flavors. (See Davies *et al.,* Recent Advances in the Generation of Chiral Intermediates Using Enzymes, CRC Press, Boca Raton, Florida (1990)). The following reactions catalyzed by enzymes are of interest to organic chemists:hydrolysis of carboxylic acid esters, phosphate esters, amides and nitriles, esterification reactions, trans-esterification reactions, synthesis of amides, reduction of alkanones and oxoalkanates, oxidation of alcohols to carbonyl compounds, oxidation of sulfides to sulfoxides, and carbon bond forming reactions such as the aldol reaction.

When considering the use of an enzyme encoded by one of the ORFs of the present invention for biotransformation and organic synthesis it is sometimes necessary to consider the respective advantages and disadvantages of using a microorganism as opposed to an isolated enzyme. Pros and cons of using a whole cell system on the one hand or an isolated partially purified enzyme on the other hand, has been described in detail by Bud *et al.,* Chemistry in Britain (1987), p. 127.

Amino transferases, enzymes involved in the biosynthesis and metabolism of amino acids, are useful in the catalytic production of amino acids. The advantages of using microbial based enzyme systems is that the amino transferase enzymes catalyze the stereo- selective synthesis of only L-amino acids and generally possess uniformly high catalytic rates. A description of the use of amino transferases for amino acid production is provided by Roselle-David, *Methods of Enzymology* 136:479 (1987).

Another category of useful proteins encoded by the ORFs of the present invention include enzymes involved in nucleic acid synthesis, repair, and recombination. A variety of commercially important enzymes have previously been isolated from members of *Staphylococcus aureus*. These include Sau3A and Sau96I.

2. Generation of Antibodies

As described here, the proteins of the present invention, as well as homologs thereof, can be used in a variety procedures and methods known in the art which are currently applied to other proteins. The proteins of the present invention can further be used to generate an antibody which selectively binds the protein. Such antibodies can be either monoclonal or polyclonal antibodies, as well fragments of these antibodies, and humanized forms.

The invention further provides antibodies which selectively bind to one of the proteins of the present invention and hybridomas which produce these antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A. M., MONOCLONAL ANTIBODY TECHNOLOGY: LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth *et al., J. Immunol. Methods* 35: 1-21 (1980), Kohler and Milstein, *Nature* 256: 495-497 (1975)), the trioma technique, the human B- cell hybridoma technique (Kozbor *et al., Immunology Today*

4: 72 (1983), pgs. 77-96 of Cole *et al.,* in MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc. (1985)).

Any animal (mouse, rabbit, *etc.*) which is known to produce antibodies can be immunized with the pseudogene polypeptide. Methods for immunization are well known in the art. Such methods include subcutaneous or interperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the protein encoded by the ORF of the present invention used for immunization will vary based on the animal which is immunized, the antigenicity of the peptide and the site of injection.

The protein which is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to coupling the antigen with a heterologous protein (such as globulin or galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells.

Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz *et al., Exp. Cell Res.* 175: 109-124 (1988)).

Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, A. M., Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)).

Techniques described for the production of single chain antibodies (U. S. Patent 4,946,778) can be adapted to produce single chain antibodies to proteins of the present invention.

For polyclonal antibodies, antibody containing antisera is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures.

The present invention further provides the above- described antibodies in detectably labelled form. Antibodies can be detectably labelled through the use of radioisotopes, affinity labels (such as biotin, avidin, *etc.*), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, *etc.*) fluorescent labels (such as FITC or rhodamine, *etc.*), paramagnetic atoms, *etc.* Procedures for accomplishing such labelling are well-known in the art, for example see Sternberger *et al.,* J. Histochem. Cytochem. 18:315 (1970); Bayer, E. A. *et al.,* Meth. Enzym. 62:308 (1979); Engval, E. *et al.,* Immunol. 109:129 (1972); Goding, J. W. J. Immunol. Meth. 13:215 (1976)).

The labeled antibodies of the present invention can be used for *in vitro, in vivo,* and in situ assays to identify cells or tissues in which a fragment of the *Staphylococcus aureus* genome is expressed.

The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D. M. *et al.,* "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W. D. *et al.,* Meth. Enzym. 34 Academic Press, N. Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro, in vivo,* and in situ assays as well as for immunoaffinity purification of the proteins of the present invention.

3. Diagnostic Assays and Kits

The present invention further provides methods to identify the expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using one of the DFs,antigens or antibodies of the present invention.

In detail, such methods comprise incubating a test sample with one or more of the antibodies, or one or more of the DFs, or one or more antigens of the present invention and assaying for binding of the DFs, antigens or antibodies to components within the test sample.

Conditions for incubating a DF, antigen or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the DF or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the Dfs, antigens or antibodies of the present invention. Examples of such assays can be found in Chard, T., An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.,* Techniques in Immunocytochemistry, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry; PCT publication WO95/32291, and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985), all of which are hereby incorporated herein by reference.

The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based

on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises:(a) a first container comprising one of the Dfs, antigens or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following:wash reagents, reagents capable of detecting presence of a bound DF, antigen or antibody.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, *etc.*), and containers which contain the reagents used to detect the bound antibody, antigen or DF.

Types of detection reagents include labelled nucleic acid probes, labelled secondary antibodies, or in the alternative, if the primary antibody is labelled, the enzymatic, or antibody binding reagents which are capable of reacting with the labelled antibody. One skilled in the art will readily recognize that the disclosed Dfs, antigens and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

4. Screening Assay for Binding Agents

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents which bind to a protein encoded by one of the ORFs of the present invention or to one of the fragments and the *Staphylococcus aureus* fragment and contigs herein described.

In general, such methods comprise steps of:

(a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention, or an isolated fragment of the Staphylococcus aureus genome; and

(b) determining whether the agent binds to said protein or said fragment.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention.

Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides, for example see Hurby *et al.*, Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W. H. Freeman, NY (1992), pp. 289-307, and Kaspczak *et al.*, Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control.

One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods usually contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee *et al.*, Nucl. Acids Res. 6:3073 (1979); Cooney *et al.*, Science 241:456 (1988); and Dervan *et al.*, Science 251: 1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated

to be effective in model systems. Information contained in the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides, and other DNA binding agents.

5. Pharmaceutical Compositions and Vaccines

The present invention further provides pharmaceutical agents which can be used to modulate the growth or pathogenicity of *Staphylococcus aureus,* or another related organism, *in vivo* or *in vitro.* As used herein, a "pharmaceutical agent" is defined as a composition of matter which can be formulated using known techniques to provide a pharmaceutical compositions. As used herein, the "pharmaceutical agents of the present invention" refers the pharmaceutical agents which are derived from the proteins encoded by the ORFs of the present invention or are agents which are identified using the herein described assays.

As used herein, a pharmaceutical agent is said to "modulate the growth or pathogenicity of *Staphylococcus aureus* or a related organism, *in vivo* or *in vitro,"* when the agent reduces the rate of growth, rate of division, or viability of the organism in question. The pharmaceutical agents of the present invention can modulate the growth or pathogenicity of an organism in many fashions, although an understanding of the underlying mechanism of action is not needed to practice the use of the pharmaceutical agents of the present invention. Some agents will modulate the growth or pathogenicity by binding to an important protein thus blocking the biological activity of the protein, while other agents may bind to a component of the outer surface of the organism blocking attachment or rendering the organism more prone to act the bodies nature immune system. Alternatively, the agent may comprise a protein encoded by one of the ORFs of the present invention and serve as a vaccine. The development and use of vaccines derived from membrane associated polypeptides are well known in the art. The inventors have identified particularly preferred immunogenic Staphylococcus aureus polypeptides for use as vaccines. Such immunogenic polypeptides are described above and summarized in Table 4, below.

As used herein, a "related organism" is a broad term which refers to any organism whose growth or pathogenicity can be modulated by one of the pharmaceutical agents of the present invention. In general, such an organism will contain a homolog of the protein which is the target of the pharmaceutical agent or the protein used as a vaccine. As such, related organisms do not need to be bacterial but may be fungal or viral pathogens.

The pharmaceutical agents and compositions of the present invention may be administered in a convenient manner, such as by the oral, topical, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes. The pharmaceutical compositions are administered in an amount which is effective for treating and/or prophylaxis of the specific indication. In general, they are administered in an amount of at least about 1 mg/kg body weight and in most cases they will be administered in an amount not in excess of about 1 g/kg body weight per day. In most cases, the dosage is from about 0.1 mg/kg to about 10 g/kg body weight daily, taking into account the routes of administration, symptoms, *etc.*

The agents of the present invention can be used in native form or can be modified to form a chemical derivative. As used herein, a molecule is said to be a "chemical derivative" of another molecule when it contains additional chemical moieties not normally a part of the molecule. Such moieties may improve the molecule's solubility, absorption, biological half life, *etc.* The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, *etc.* Moieties capable of mediating such effects are disclosed in, among other sources, REMINGTON'S PHARMACEUTICAL SCIENCES (1980) cited elsewhere herein.

For example, such moieties may change an immunological character of the functional derivative, such as affinity for a given antibody. Such changes in immunomodulation activity are measured by the appropriate assay, such as a competitive type immunoassay. Modifications of such protein properties as redox or thermal stability, biological half-life, hydrophobicity, susceptibility to proteolytic degradation or the tendency to aggregate with carriers or into multimers also may be effected in this way and can be assayed by methods well known to the skilled artisan.

The therapeutic effects of the agents of the present invention may be obtained by providing the agent to a patient by any suitable means (*e.g.,* inhalation, intravenously, intramuscularly, subcutaneously, enterally, or parenterally). It is preferred to administer the agent of the present invention so as to achieve an effective concentration within the blood or tissue in which the growth of the organism is to be controlled. To achieve an effective blood concentration, the preferred method is to administer the agent by injection. The administration may be by continuous infusion, or by single or multiple injections.

In providing a patient with one of the agents of the present invention, the dosage of the administered agent will vary depending upon such factors as the patient's age, weight, height, sex, general medical condition, previous medical history, *etc.* In general, it is desirable to provide the recipient with a dosage of agent which is in the range of from about 1 pg/kg to 10 mg/kg (body weight of patient), although a lower or higher dosage may be administered. The therapeutically effective dose can be lowered by using combinations of the agents of the present invention or another agent.

As used herein, two or more compounds or agents are said to be administered "in combination" with each other when either (1) the physiological effects of each compound, or (2) the serum concentrations of each compound can

be measured at the same time. The composition of the present invention can be administered concurrently with, prior to, or following the administration of the other agent.

The agents of the present invention are intended to be provided to recipient subjects in an amount sufficient to decrease the rate of growth (as defined above) of the target organism.

The administration of the agent(s) of the invention may be for either a "prophylactic" or "therapeutic" purpose. When provided prophylactically, the agent(s) are provided in advance of any symptoms indicative of the organisms growth. The prophylactic administration of the agent(s) serves to prevent, attenuate, or decrease the rate of onset of any subsequent infection. When provided therapeutically, the agent(s) are provided at (or shortly after) the onset of an indication of infection. The therapeutic administration of the compound(s) serves to attenuate the pathological symptoms of the infection and to increase the rate of recovery.

The agents of the present invention are administered to a subject, such as a mammal, or a patient, in a pharmaceutically acceptable form and in a therapeutically effective concentration. A composition is said to be "pharmacologically acceptable" if its administration can be tolerated by a recipient patient. Such an agent is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient patient.

The agents of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby these materials, or their functional derivatives, are combined in admixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation, inclusive of other human proteins, *e. g.*, human serum albumin, are described, for example, in REMINGTON'S PHARMACEUTICAL SCIENCES, 16th Ed., Osol, A., Ed., Mack Publishing, Easton PA (1980). In order to form a pharmaceutically acceptable composition suitable for effective administration, such compositions will contain an effective amount of one or more of the agents of the present invention, together with a suitable amount of carrier vehicle.

Additional pharmaceutical methods may be employed to control the duration of action. Control release preparations may be achieved through the use of polymers to complex or absorb one or more of the agents of the present invention. The controlled delivery may be effectuated by a variety of well known techniques, including formulation with macromolecules such as, for example, polyesters, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or protamine, sulfate, adjusting the concentration of the macromolecules and the agent in the formulation, and by appropriate use of methods of incorporation, which can be manipulated to effectuate a desired time course of release. Another possible method to control the duration of action by controlled release preparations is to incorporate agents of the present invention into particles of a polymeric material such as polyesters, polyamino acids, hydrogels, poly(lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is possible to entrap these materials in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization with, for example, hydroxymethylcellulose or gelatine-microcapsules and poly(methylmethacylate) microcapsules, respectively, or in colloidal drug delivery systems, for example, liposomes, albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions. Such techniques are disclosed in REMINGTON'S PHARMACEUTICAL SCIENCES (1980).

The invention further provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

In addition, the agents of the present invention may be employed in conjunction with other therapeutic compounds.

6. Shot-Gun Approach to Megabase DNA Sequencing

The present invention further demonstrates that a large sequence can be sequenced using a random shotgun approach. This procedure, described in detail in the examples that follow, has eliminated the up front cost of isolating and ordering overlapping or contiguous subclones prior to the start of the sequencing protocols.

Certain aspects of the present invention are described in greater detail in the examples that follow. The examples are provided by way of illustration. Other aspects and embodiments of the present invention are contemplated by the inventors, as will be clear to those of skill in the art from reading the present disclosure.

ILLUSTRATIVE EXAMPLES

LIBRARIES AND SEQUENCING

1. Shotgun Sequencing Probability Analysis

The overall strategy for a shotgun approach to whole genome sequencing follows from the Lander and Waterman (Landerman and Waterman, Genomics 2: 231 (1988)) application of the equation for the Poisson distribution. According to this treatment, the probability, P_0, that any given base in a sequence of size L, in nucleotides, is not sequenced after a certain amount, n, in nucleotides, of random sequence has been determined can be calculated by the equation $P_0 = e^{-m}$, where m is L/n, the fold coverage." For instance, for a genome of 2.8 Mb, m=1 when 2.8 Mb of sequence has been randomly generated (1X coverage). At that point, $P_0 = e^{-1} = 0.37$. The probability that any given base has not been sequenced is the same as the probability that any region of the whole sequence L has not been determined and, therefore, is equivilent to the fraction of the whole sequence that has yet to be determined. Thus, at one-fold coverage, approximately 37% of a polynucleotide of size L, in nucleotides has not been sequenced. When 14 Mb of sequence has been generated, coverage is 5X for a .2.8 Mb and the unsequenced fraction drops to .0067 or 0.67%. 5X coverage of a 2.8 Mb sequence can be attained by sequencing approximately 17,000 random clones from both insert ends with an average sequence read length of 410 bp.

Similarly, the total gap length, G, is determined by the equation $G = Le^{-m}$, and the average gap size, g, follows the equation, g = L/n. Thus, 5X coverage leaves about 240 gaps averaging about 82 bp in size in a sequence of a polynucleotide 2.8 Mb long.

The treatment above is essentially that of Lander and Waterman, *Genomics* 2: 231 (1988).

2. Random Library Construction

In order to approximate the random model described above during actual sequencing, a nearly ideal library of cloned genomic fragments is required. The following library construction procedure was developed to achieve this end.

Staphylococcus aureus DNA was prepared by phenol extraction. A mixture containing 600 ug DNA in 3.3 ml of 300 mM sodium acetate, 10 mM Tris-HCl, 1 mM Na-EDTA, 30% glycerol was sonicated for 1 min. at 0°C in a Branson Model 450 Sonicator at the lowest energy setting using a 3 mm probe. The sonicated DNA was ethanol precipitated and redissolved in 500 ul TE buffer.

To create blunt-ends, a 100 ul aliquot of the resuspended DNA was digested with 5 units of BAL31 nuclease (New England BioLabs) for 10 min at 30°C in 200 ul BAL31 buffer . The digested DNA was phenol-extracted, ethanol-precipitated, redissolved in 100 ul TE buffer, and then size-fractionated by electrophoresis through a 1.0% low melting temperature agarose gel. The section containing DNA fragments 1.6-2.0 kb in size was excised from the gel, and the LGT agarose was melted and the resulting solution was extracted with phenol to separate the agarose from the DNA. DNA was ethanol precipitated and redissolved in 20 ul of TE buffer for ligation to vector.

A two-step ligation procedure was used to produce a plasmid library with 97% inserts, of which >99% were single inserts. The first ligation mixture (50 ul) contained 2 ug of DNA fragments, 2 ug pUC18 DNA (Pharmacia) cut with SmaI and dephosphorylated with bacterial alkaline phosphatase, and 10 units of T4 ligase (GIBCO/BRL) and was incubated at 14°C for 4 hr. The ligation mixture then was phenol extracted and ethanol precipitated, and the precipitated DNA was dissolved in 20 ul TE buffer and electrophoresed on a 1.0% low melting agarose gel. Discrete bands in a ladder were visualized by ethidium bromide-staining and UV illumination and identified by size as insert (i), vector (v), v+i, v+2i, v+3i, *etc.* The portion of the gel containing v+i DNA was excised and the v+i DNA was recovered and resuspended into 20 ul TE. The v+i DNA then was blunt-ended by T4 polymerase treatment for 5 min. at 37° C in a reaction mixture (50 ul) containing the v+i linears, 500 uM each of the 4 dNTPs, and 9 units of T4 polymerase (New England BioLabs), under recommended buffer conditions. After phenol extraction and ethanol precipitation the repaired v+i linears were dissolved in 20 ul TE. The final ligation to produce circles was carried out in a 50 ul reaction containing 5 ul of v+i linears and 5 units of T4 ligase at 14°C overnight. After 10 min. at 70°C the following day, the reaction mixture was stored at -20°C.

This two-stage procedure resulted in a molecularly random collection of single-insert plasmid recombinants with minimal contamination from double-insert chimeras (<1%) or free vector (<3%).

Since deviation from randomness can arise from propagation the DNA in the host, *E.coli* host cells deficient in all recombination and restriction functions (A. Greener, Strategies 3 (1):5 (1990)) were used to prevent rearrangements, deletions, and loss of clones by restriction. Furthermore, transformed cells were plated directly on antibiotic diffusion plates to avoid the usual broth recovery phase which allows multiplication and selection of the most rapidly growing cells.

Plating was carried out as follows. A 100 ul aliquot of Epicurian Coli SURE II Supercompetent Cells (Stratagene 200152) was thawed on ice and transferred to a chilled Falcon 2059 tube on ice. A 1.7 ul aliquot of 1.42 M beta-

mercaptoethanol was added to the aliquot of cells to a final concentration of 25 mM. Cells were incubated on ice for 10 min. A 1 ul aliquot of the final ligation was added to the cells and incubated on ice for 30 min. The cells were heat pulsed for 30 sec. at 42° C and placed back on ice for 2 min. The outgrowth period in liquid culture was eliminated from this protocol in order to minimize the preferential growth of any given transformed cell. Instead the transformation mixture was plated directly on a nutrient rich SOB plate containing a 5 ml bottom layer of SOB agar (5% SOB agar: 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 1.5% Difco Agar per liter of media). The 5 ml bottom layer is supplemented with 0.4 ml of 50 mg/ml ampicillin per 100 ml SOB agar. The 15 ml top layer of SOB agar is supplemented with 1 ml X-Gal (2%), 1 ml $MgCl_2$ (1 M), and 1 ml $MgSO_4$/100 ml SOB agar. The 15 ml top layer was poured just prior to plating. Our titer was approximately 100 colonies/10 ul aliquot of transformation.

All colonies were picked for template preparation regardless of size. Thus, only clones lost due to "poison" DNA or deleterious gene products would be deleted from the library, resulting in a slight increase in gap number over that expected.

3. Random DNA Sequencing

High quality double stranded DNA plasmid templates were prepared using an alkaline lysis method developed in collaboration with 5Prime → 3Prime Inc. (Boulder, CO). Plasmid preparation was performed in a 96-well format for all stages of DNA preparation from bacterial growth through final DNA purification. Average template concentration was determined by running 25% of the samples on an agarose gel. DNA concentrations were not adjusted.

Templates were also prepared from a *Staphylococcus aureus* lambda genomic library. An unamplified library was constructed in Lambda DASH II vector (Stratagene). *Staphylococcus aureus* DNA (> 100 kb) was partially digested in a reaction mixture (200 ul) containing 50 ug DNA, 1X Sau3AI buffer, 20 units Sau3AI for 6 min. at 23 C. The digested DNA was phenol-extracted and centrifuges over a 10- 40% sucroce gradient. Fractions containing genomic DNA of 15-25 kb were recovered by precipitation . One ul of fragments was used with 1 ul of DASHII vector (Stratagene) in the recommended ligation reaction. One ul of the ligation mixture was used per packaging reaction following the recommended protocol with the Gigapack II XL Packaging Extract Phage were plated directly without amplification from the packaging mixture (after dilution with 500 ul of recommended SM buffer and chloroform treatment). Yield was about $2.5x10^9$ pfu/ul.

An amplified library was prepared from the primary packaging mixture according to the manufactureer's protocol. The amplified library is stored frozen in 7% dimethylsulfoxide. The phage titer is approximately $1x10^9$ pfu/ml.

Mini-liquid lysates (0.1ul) are prepared from randomly selected plaques and template is prepared by long range PCR. Samples are PCR amplified using modified T3 and T7 primers, and Elongase Supermix (LTI).

Sequencing reactions are carried out on plasmid templates using a combination of two workstations (BIOMEK 1000 and Hamilton Microlab 2200) and the Perkin-Elmer 9600 thermocycler with Applied Biosystems PRISM Ready Reaction Dye Primer Cycle Sequencing Kits for the M13 forward (M13-21) and the M13 reverse (M13RP1) primers. Dye terminator sequencing reactions are carried out on the lambda templates on a Perkin-Elmer 9600 Thermocycler using the Applied Biosystems Ready Reaction Dye Terminator Cycle Sequencing kits. Modified T7 and T3 primers are used to sequence the ends of the inserts from the Lambda DASH II library. Sequencing reactions are on a combination of AB 373 DNA Sequencers and ABI 377 DNA sequencers. All of the dye terminator sequencing reactions are analyzed using the 2X 9 hour module on the AB 377. Dye primer reactions are analyzed on a combination of ABI 373 and ABI 377 DNA sequencers. The overall sequencing success rate very approximately is about 85% for M13-21 and M13RP1 sequences and 65% for dye-terminator reactions. The average usable read length is 485 bp for M13-21 sequences, 445bp for M13RP1 sequences, and 375 bp for dye-terminator reactions.

4. Protocol for Automated Cycle Sequencing

The sequencing was carried out using Hamilton Microstation 2200, Perkin Elmer 9600 thermocyclers, ABI 373 and ABI 377 Automated DNA Sequencers. The Hamilton combines pre-aliquoted templates and reaction mixes consisting of deoxy- and dideoxynucleotides, the thermostable Taq DNA polymerase, fluorescently-labelled sequencing primers, and reaction buffer. Reaction mixes and templates were combined in the wells of a 96-well thermocycling plate and transferred to the Perkin Elmer 9600 thermocycler. Thirty consecutive cycles of linear amplification (i.e.., one primer synthesis) steps were performed including denaturation, annealing of primer and template, and extension; i.e., DNA synthesis. A heated lid with rubber gaskets on the thermocycling plate prevents evaporation without the need for an oil overlay.

Two sequencing protocols were used: one for dye-labelled primers and a second for dye-labelled dideoxy chain terminators. The shotgun sequencing involves use of four dye-labelled sequencing primers, one for each of the four terminator nucleotide. Each dye-primer was labelled with a different fluorescent dye, permitting the four individual reactions to be combined into one lane of the 373 or 377 DNA Sequencer for electrophoresis, detection, and base-

calling. ABI currently supplies premixed reaction mixes in bulk packages containing all the necessary non-template reagents for sequencing. Sequencing can be done with both plasmid and PCR-generated templates with both dye-primers and dye- terminators with approximately equal fidelity, although plasmid templates generally give longer usable sequences.

Thirty-two reactions were loaded per ABI 373 Sequencer each day and 96 samples can be loaded on an ABI 377 per day. Electrophoresis was run overnight (ABI 373) or for 2 1/2 hours (ABI 377) following the manufacturer's protocols. Following electrophoresis and fluorescence detection, the ABI 373 or ABI 377 performs automatic lane tracking and base-calling. The lane-tracking was confirmed visually. Each sequence electropherogram (or fluorescence lane trace) was inspected visually and assessed for quality. Trailing sequences of low quality were removed and the sequence itself was loaded via software to a Sybase database (archived daily to 8mm tape). Leading vector polylinker sequence was removed automatically by a software program. Average edited lengths of sequences from the standard ABI 373 or ABI 377 were around 400 bp and depend mostly on the quality of the template used for the sequencing reaction.

INFORMATICS

1. Data Management

A number of information management systems for a large-scale sequencing lab have been developed. (For review see, for instance, Kerlavage *et al., Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences,* IEEE Computer Society Press, Washington D. C., 585 (1993)) The system used to collect and assemble the sequence data was developed using the Sybase relational database management system and was designed to automate data flow whereever possible and to reduce user error. The database stores and correlates all information collected during the entire operation from template preparation to final analysis of the genome. Because the raw output of the ABI 373 Sequencers was based on a Macintosh platform and the data management system chosen was based on a Unix platform, it was necessary to design and implement a variety of multi- user, client-server applications which allow the raw data as well as analysis results to flow seamlessly into the database with a minimum of user effort.

2. Assembly

An assembly engine (TIGR Assembler) developed for the rapid and accurate assembly of thousands of sequence fragments was enployed to generate contigs. The TIGR assembler simultaneously clusters and assembles fragments of the genome. In order to obtain the speed necessary to assemble more than 10^4 fragments, the algorithm builds a hash table of 12 bp oligonucleotide subsequences to generate a list of potential sequence fragment overlaps. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Beginning with a single seed sequence fragment, TIGR Assembler extends the current contig by attempting to add the best matching fragment based on oligonucleotide content. The contig and candidate fragment are aligned using a modified version of the Smith-Waterman algorithm which provides for optimal gapped alignments (Waterman, M. S., *Methods in Enzymology* 164: 765 (1988)). The contig is extended by the fragment only if strict criteria for the quality of the match are met. The match criteria include the minimum length of overlap, the maximum length of an unmatched end, and the minimum percentage match. These criteria are automatically lowered by the algorithm in regions of minimal coverage and raised in regions with a possible repetitive element. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Fragments representing the boundaries of repetitive elements and potentially chimeric fragments are often rejected based on partial mismatches at the ends of alignments and excluded from the current contig. TIGR Assembler is designed to take advantage of clone size information coupled with sequencing from both ends of each template. It enforces the constraint that sequence fragments from two ends of the same template point toward one another in the contig and are located within a certain ranged of base pairs (definable for each clone based on the known clone size range for a given library).

3. Identifying Genes

The predicted coding regions of the *Staphylococcus aureus* genome were initially defined with the program zorf, which finds ORFs of a minimum length. The predicted coding region sequences were used in searches against a database of all *Staphylococcus aureus* nucleotide sequences from GenBank (release 92.0), using the BLASTN search method to identify overlaps of 50 or more nucleotides with at least a 95% identity. Those ORFs with nucleotide sequence matches are shown in Table 1. The ORFs without such matches were translated to protein sequences and and compared to a non-redundant database of known proteins generated by combining the Swiss-prot, PIR and GenPept databases. ORFs of at least 80 amino acids that matched a database protein with BLASTP probability less than or equal to 0.01 are shown in Table 2. The table also lists assigned functions based on the closest match in the databases.

ORFs of at least 120 amino acids that did not match protein or nucleotide sequences in the databases at these levels are shown in Table 3.

ILLUSTRATIVE APPLICATIONS

1. Production of an Antibody to a *Staphylococcus aureus* Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells using any one of the methods known in the art. The protein can also be produced in a recombinant prokaryotic expression system, such as *E. coli*, or can by chemically synthesized. Concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows.

2. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., Nature 256:495 (1975) or modifications of the methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., Meth. Enzymol. 70:419 (1980), and modified methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al.* Basic Methods in Molecular Biology Elsevier, New York. Section 21-2 (1989).

3. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than other and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigenadministered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. *et al.*, J. Clin. Endocrinol. Metab. 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. *et al.*, Chap. 19 in:Handbook of Experimental Immunology, Wier, D., ed, Blackwell (1973). Plateau concentration of antibody is usually in the range of 0. 1 to 0. 2 mg/ml of serum (about 12M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in:Manual of Clinical Immunology, second edition, Rose and Friedman, eds., Amer. Soc. For Microbiology, Washington, D. C. (1980)

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi- quantitatively or qualitatively to identify the presence of antigen in a biological sample. In addition, they are useful in various animal models of Staphylococcal disease known to those of skill in the art as a means of evaluating the protein used to make the antibody as a potential vaccine target or as a means of evaluating the antibody as a potential immunothereapeutic reagent.

3. Preparation of PCR Primers and Amplification of DNA

Various fragments of the *Staphylococcus aureus* genome, such as those of Tables 1-3 and SEQ ID NOS:1-5,191 can be used, in accordance with the present invention, to prepare PCR primers for a variety of uses. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. When selecting a primer sequence, it is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approxi-

mately the same. The PCR primers and amplified DNA of this Example find use in the Examples that follow.

4. Gene expression from DNA Sequences Corresponding to ORFs

A fragment of the *Staphylococcus aureus* genome provided in Tables 1-3 is introduced into an expression vector using conventional technology. Techniques to transfer cloned sequences into expression vectors that direct protein translation in mammalian, yeast, insect or bacterial expression systems are well known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism, as explained by Hatfield *et al.,* U. S. Patent No. 5,082,767, incorporated herein by this reference.

The following is provided as one exemplary method to generate polypeptide(s) from cloned ORFs of the *Staphylococcus aureus* genome fragment. Bacterial ORFs generally lack a poly A addition signal. The addition signal sequence can be added to the construct by, for example, splicing out the poly A addition sequence from pSG5 (Stratagene) using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene) for use in eukaryotic expression systems. pXT1 contains the LTRs and a portion of the gag gene of Moloney Murine Leukemia Virus. The positions of theLTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex thymidine kinase promoter and the selectable neomycin gene. The *Staphylococcus aureus* DNA is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the *Staphylococcus aureus* DNA and containing restriction endonuclease sequences for PstI incorporated into the 5' primer and BglII at the 5' end of the corresponding *Staphylococcus aureus* DNA 3' primer, taking care to ensure that the *Staphylococcus aureus* DNA is positioned such that its followed with the poly A addition sequence. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with BglII, purified and ligated to pXT1, now containing a poly A addition sequence and digested BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600 ug/ml G418 (Sigma, St. Louis, Missouri). The protein is preferably released into the supernatant. However if the protein has membrane binding domains, the protein may additionally be retained within the cell or expression may be restricted to the cell surface. Since it may be necessary to purify and locate the transfected product, synthetic 15-mer peptides synthesized from the predicted *Staphylococcus aureus* DNA sequence are injected into mice to generate antibody to the polypeptide encoded by the *Staphylococcus aureus* DNA.

Alternativly and if antibody production is not possible, the *Staphylococcus aureus* DNA sequence is additionally incorporated into eukaryotic expression vectors and expressed as, for example, a globin fusion. Antibody to the globin moiety then is used to purify the chimeric protein. Corresponding protease cleavage sites are engineered between the globin moiety and the polypeptide encoded by the *Staphylococcus aureus* DNA so that the latter may be freed from the formed by simple protease digestion. One useful expression vector for generating globin chimerics is pSG5 (Stratagene). This vector encodes a rabbit globin. Intron II of the rabbit globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.,* cited elsewhere herein, and many of the methods are available from the technical assistance representatives from Stratagene, Life Technologies, Inc., or Promega. Polypeptides of the invention also may be produced using *in vitro* translation systems such as *in vitro* ExpressTM Translation Kit (Stratagene).

While the present invention has been described in some detail for purposes of clarity and understanding, one skilled in the art will appreciate that various changes in form and detail can be made without departing from the true scope of the invention.

All patents, patent applications and publications referred to above are hereby incorporated by reference.

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name ( )	percent ident	HSP nt length	ORF nt length
1	1	1419	757	emb|X17301|SAHD	S.aureus DNA for hld gene and for part of agr gene	100	663	663
1	2	3273	2452	emb|X52543|SAAG	S.aureus agrA, agrB and hld genes	99	809	822
1	5	6418	5651	dbj|D14711|STAH	Staphylococcus aureus HSP10 and HSP60 genes	98	223	768
5	1	807	439	emb|X72700|SAPV	S.aureus genes for S and F components of Panton-Valentine leucocidins	81	216	369
5	4	5031	3571	emb|X72700|SAPV	S.aureus genes for S and F components of Panton-Valentine leucocidins	95	424	1461
10	1	86	904	gb|L25288|	Staphylococcus aureus gyrase-like protein alpha and beta subunit (grlA and grlB) genes, complete cds	98	715	819
16	5	5302	6246	gb|U35773|	Staphylococcus aureus prolipoprotein diacylglyceryl transferase (lgt) gene, complete cds	94	251	945
16	6	6249	7091	gb|U35773|	Staphylococcus aureus prolipoprotein diacylglyceryl transferase (lgt) gene, complete cds	99	843	843
16	7	7084	7584	gb|U35773|	Staphylococcus aureus prolipoprotein diacylglyceryl transferase (lgt) gene, complete cds	99	342	501
20	1	995	549	gb|L19300|	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	100	443	447
20	2	1011	841	gb|L19300|	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	91	137	171
20	3	2010	1798	gb|L19300|	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	100	110	213
20	4	5300	3825	gb|M76714|	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	100	948	1476
20	5	4788	4282	gb|M76714|	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	100	309	507
26	1	2	145	gb|U41072|	Staphylococcus aureus isoleucyl-tRNA synthetase (ileS) gene, partial cds	100	126	144
26	2	84	557	gb|U41072|	Staphylococcus aureus isoleucyl-tRNA synthetase (ileS) gene, partial cds	99	430	474
26	3	763	3531	emb|X74219|SAIL	S.aureus gene for isoleucyl-tRNA synthetase	99	2769	2769
29	3	1261	4392	gb|U66665|	Staphylococcus aureus DNA fragment with class II promoter activity	100	117	3132
31	14	14977	13463	emb|X73889|SAP1	S.aureus genes P1 and P2	99	1351	1515
31	15	14241	13855	emb|X73889|SAP1	S.aureus genes P1 and P2	98	258	387
38	17	14284	13112	gb|M12715|	S.aureus geh gene encoding lipase (glycerol ester hydrolase)	100	372	1173
38	19	13434	15518	gb|M12715|	S.aureus geh gene encoding lipase (glycerol ester hydrolase)	100	2085	2085

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
46	2	519	1727	gb|U73374|	Staphylococcus aureus type 8 capsulo genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1209	1209
46	3	1720	2295	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	576	576
46	4	2259	3182	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	97	924	924
46	5	3173	4498	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1283	1326
46	6	4536	5720	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1185	1185
46	7	6455	6120	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	278	336
48	1	2	955	gb|L25893|	Staphylococcus aureus recA gene, complete cds	99	954	954
50	3	4465	2924	emb|X85029|SAAH	S.aureus AhpC gene	100	88	1542
50	4	4108	3515	emb|X85029|SAAH	S.aureus AhpC gene	98	540	594
54	3	5074	3392	emb|X62992|SAFN	S.aureus fnbB gene for fibronectin binding protein B	100	1668	1683
54	4	4865	4122	emb|X62992|SAFN	S.aureus fnbB gene for fibronectin binding protein B	99	720	744
54	5	5056	4562	emb|X62992|SAFN	S.aureus fnbB gene for fibronectin binding protein B	100	463	495
54	6	11386	8300	gb|J04151|	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	100	3087	3087
58	3	1743	2819	emb|X87104|SADN	S.aureus mdr, pbp4 and taqD genes (SG511-55 isolate)	89	68	1077
58	4	2858	3280	emb|X91786|SAPB	S.aureus abcA, pbp4, and tagD genes	99	423	423
58	5	6005	4701	emb|X91786|SAPB	S.aureus abcA, pbp4, and tagD genes	99	1305	1305
58	6	5677	5378	gb|U29478|	Staphylococcus aureus ABC transporter-like protein AbcA (abcA) gene, partial cds	100	300	300
58	7	5086	6840,	emb|X91786|SAPB	S.aureus abcA, pbp4, and tagD genes	99	1755	1755
72	1	888	445	gb|U21854|	S.aureus agr gene encoding an accessory gene regulator protein, complete cds	100	444	444
72	2	2457	1453	emb|X52543|SAAG	S.aureus agrA, agrB and hld genes	99	673	1005

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
82	1	357	3917	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	2396	3561
82	2	4027	7677	emb|X89233|SARP	S.aureus DNA for rpoC gene	99	3171	3651
82	3	7745	8068	gb|U20869|	Staphylococcus aureus ribosomal protein S12 (rpsL) gene, complete cds, ribosomal protein S7 (rpsG) and ORF 1 genes, partial cds	100	320	324
82	4	8103	8579	gb|U20869|	Staphylococcus aureus ribosomal protein S12 (rpsL) gene, complete cds, ribosomal protein S7 (rpsG) and ORF 1 genes, partial cds	100	477	477
82	5	8618	8821	gb|U20869|	Staphylococcus aureus ribosomal protein S12 (rpsL) gene, complete cds, ribosomal protein S7 (rpsG) and ORF 1 genes, partial cds	100	154	204
84	1	18	191	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	164	174
84	2	189	893	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	94	705	705
84	3	887	1660	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	774	774
84	4	1584	3503	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1920	1920
84	5	3394	4521	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	97	1128	1128
84	6	4519	5643	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	97	1125	1125
96	2	1245	3896	emb|Z18852|SACF	S.aureus gene for clumping factor	83	660	2652
97	2	625	882	gb|U41072|	Staphylococcus aureus isoleucyl-tRNA synthetase (ileS) gene, partial cds	97	68	258
111	1	3	452	gb|L41499|	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	100	450	450
111	2	526	1041	gb|L41499|	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	99	516	516
117	2	1278	1958	gb|M83994|	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	100	61	681

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
118	4	3787	4254	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	99	467	468
130	4	2597	3640	emb|X13290|SATN	Staphylococcus aureus multi-resistance plasmid pSK1 DNA containing transposon Tn4003	78	956	1044
130	5	3813	4265	emb|Z16422|SADI	S.aureus dfrB gene for dihydrofolate reductase	98	416	453
130	6	4309	5172	emb|Z16422|SADI	S.aureus dfrB gene for dihydrofolate reductase	98	607	864
136	4	5296	6207	emb|X71437|SAGY	S.aureus genes gyrB, gyrA and recF (partial)	97	838	912
136	5	11680	8987	dbj|D10489|STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	100	2694	2694
136	6	12886	10940	dbj|D10489|STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	99	1947	1947
136	7	12592	11765	gb|S77055|	recF cluster: dnaA=replisome assembly protein...gyrB=DNA gyrase beta subunit [Staphylococcus aureus, YB886, Genomic, 5 genes, 3573 nt]	99	822	828
143	3	4171	2867	gb|U36379|	Staphylococcus aureus S-adenosylmethionine synthetase gene, complete cds	99	1305	1305
143	4	3100	4281	gb|L42943|	Staphylococcus aureus (clone KIN50) phosphoenolpyruvate carboxykinase (pckA) gene, complete cds	100	1170	1182
143	5	4254	4718	gb|U51133|	Staphylococcus aureus phosphoenolpyruvate carboxykinase (pcka) gene, complete cds	100	449	465
143	9	6977	7261	gb|U51132|	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (mene), and o-succinylbenzoic acid synthetase (menc) genes, complete cds	100	75	285
143	10	9464	8361	gb|U51132|	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (mene), and o-succinylbenzoic acid synthetase (menc) genes, complete cds	100	1104	1104
143	11	11232	9748	gb|U51132|	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (mene), and o-succinylbenzoic acid synthetase (menc) genes, complete cds	100	1485	1485
143	12	10739	10320	gb|U51132|	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (mene), and o-succinylbenzoic acid synthetase (menc) genes, complete cds	100	332	420
152	5	2454	3437	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	99	305	984
152	6	3513	4820	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	98	1308	1308
152	7	4818	6230	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	99	1413	1413
153	1	387	1526	gb|S77055|	recF cluster: dnaA=replisome assembly protein...gyrB=DNA gyrase beta subunit [Staphylococcus aureus, YB886, Genomic, 5 genes, 3573 nt]	99	1140	1140
153	2	1877	2152	gb|S77055|	recF cluster: dnaA=replisome assembly protein...gyrB=DNA gyrase beta subunit [Staphylococcus aureus, YB886, Genomic, 5 genes, 3573 nt]	100	276	276

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
153	3	2143	2289	gb|S77055|	recF cluster: dnaA=replisome assembly protein...gyrB=DNA gyrase beta subunit (Staphylococcus aureus, YB886, Genomic, 5 genes, 3573 nt)	99	113	147
154	10	10792	9314	gb|U06451|	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	91	154	1479
154	11	9935	9615	gb|U06451|	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	99	229	321
154	12	9943	10167	gb|U06451|	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	94	123	225
154	13	10089	11501	gb|U06451|	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	99	1326	1413
159	2	2195	1212	dbj|D28879|STAP	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	100	71	984
161	3	2596	2270	gb|M83994|	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	92	203	327
162	1	1406	705	gb|U21221|	Staphylococcus aureus hyaluronate lyase (hysA) gene, complete cds	100	702	702
163	4	1263	1772	gb|U19770|	Staphylococcus aureus pyrrolidone carboxyl peptidase (pcp) gene, complete cds	96	127	510
164	7	4774	9117	dbj|D86727|D867	Staphylococcus aureus DNA for DNA polymerase III, complete cds	99	3470	4344
168	7	7448	6447	gb|U21636|	Staphylococcus aureus cmp-binding-factor 1 (cbf1) and ORF X genes, complete cds	100	1002	1002
168	8	9538	7961	gb|U21636|	Staphylococcus aureus cmp-binding-factor 1 (cbf1) and ORF X genes, complete cds	99	1158	1578
173	6	9240	7801	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	100	1440	1440
173	7	11252	9522	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	99	1731	1731
173	8	8285	8704	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	100	420	420
173	9	10168	9839	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	100	330	330
173	10	11815	10829	emb|X14827|SALA	Staphylococcus aureus lacC and lacD genes	100	987	987
173	11	12721	11774	emb|X14827|SALA	Staphylococcus aureus lacC and lacD genes	100	948	948
173	12	12838	12305	gb|M64724|	S.aureus tagatose 6-phosphate isomerase gene, complete cds	100	534	534
173	13	13243	12773	gb|M32103|	Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA), partial cds	100	471	471
173	14	14633	13866	gb|M32103|	Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA), partial cds	100	768	768

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
178	1	2	655	gb|U52961|	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	115	654
178	2	2201	1482	gb|U52961|	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	720	720
178	3	2361	1909	gb|U52961|	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	453	453
178	4	1551	1853	gb|U52961|	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	303	303
178	5	3541	2777	gb|L42945|	Staphylococcus aureus lytS and lytR genes, complete cds	99	765	765
178	6	3294	3025	gb|L42945|	Staphylococcus aureus lytS and lytR genes, complete cds	99	270	270
181	1	1114	590	gb|M61177|	S.aureus sigma factor (plaC) gene, complete cds	99	499	525
182	1	3	341	emb|X61307|SASP	Staphylococcus aureus spa gene for protein A	98	277	339
182	2	690	2312	gb|J01786|	S.aureus spa gene coding for protein A, complete cds	97	1332	1623
182	3	5861	4251	emb|X61307|SASP	Staphylococcus aureus spa gene for protein A	99	119	1611
185	1	3	824	gb|U31979|	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydrosuinate synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	90	132	822
191	3	841	2760	emb|X17679|SACO	Staphylococcus aureus coa gene for coagulase	99	1920	1920
191	4	2967	3143	emb|X16457|SAST	Staphylococcus aureus gene for staphylocoagulase	99	177	177
191	5	5768	4566	emb|X16457|SAST	Staphylococcus aureus gene for staphylocoagulase	99	250	1203
196	1	1741	872	gb|L36472|	Staphylococcus aureus lysyl-tRNA sythetase gene, complete cds, transfer RNA (tRNA) genes, 5S ribosomal RNA (5S rRNA) gene, 16S ribosomal RNA (16S rRNA) gene, 23S ribosomal RNA (23S rRNA) gene	99	870	870
198	3	1688	2011	emb|X93205|SAPT	S.aureus ptsH and ptsI genes	99	324	324
198	4	2005	2310	emb|X93205|SAPT	S.aureus ptsH and ptsI genes	97	304	306
202	1	163	1305	emb|X97985|SA12	S.aureus orfs 1,2,3 & 4	99	1143	1143
202	2	1303	2175	emb|X73889|SAP1	S.aureus genes P1 and P2	94	444	873
210	1	3114	1558	dbj|D17366|STAA	Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	99	1552	1557
210	2	2939	2232	gb|L41499|	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	99	684	708
214	11	7429	7770	dbj|D86240|D862	Staphylococcus aureus gene for unkown function and dlt operon dltA, dltB, dltC and dltD genes,complete cds	96	157	342

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
216	3	398	1318	emb|X72700|SAPV	S.aureus genes for S and F components of Panton-Valentine leucocidins	88	265	921
219	2	1810	1073	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	100	60	738
219	3	2979	2035	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	99	945	945
219	4	4359	3196	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	99	1164	1164
219	5	7044	5176	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	98	1869	1869
219	6	6557	5883	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	99	675	675
219	7	6801	6334	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	98	468	468
221	8	10816	10034	gb|L19298|	Staphylococcus aures phosphatidylinositol-specific phospholipase C (plc) gene, complete cds	91	67	783
223	1	2855	1506	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	102	1350
234	1	2	1357	emb|X97985|SA12	S.aureus orfs 1,2,3 & 4	100	176	1356
234	2	1694	2485	emb|X97985|SA12	S.aureus orfs 1,2,3 & 4	100	792	792
234	3	2648	3148	emb|X97985|SA12	S.aureus orfs 1,2,3 & 4	99	501	501
234	4	3120	4604	emb|X97985|SA12	S.aureus orfs 1,2,3 & 4	99	1305	1485
236	6	3826	5322	gb|U48826|	Staphylococcus aureus elastin binding protein (ebpS) gene, complete cds	96	648	1497
248	1	2	403	emb|X62288|SAPE	S.aureus DNA for penicillin-binding protein 2	100	103	402
248	2	388	852	gb|L25426|	Staphylococcus aureus penicillin-binding protein 2 (pbp2) gene, complete cds	99	465	465
253	2	1539	1093	gb|U46541|	Staphylococcus aureus sarA gene, complete cds	96	447	447
254	2	150	1835	gb|U57060|	Staphylococcus aureus scdA gene, complete cds	94	142	1686
254	3	1973	2728	gb|U57060|	Staphylococcus aureus scdA gene, complete cds	99	756	756
260	1	2	1900	gb|M90693|	Staphylococcus aureus glycerol ester hydrolase (lip) gene, complete cds	99	1213	1899
265	1	1	942	dbj|D21131|STAS	Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds	99	941	942

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
265	2	688	476	dbj|D21131|STAS	Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds	99	213	213
265	3	2418	1765	dbj|D21131|STAS	Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds	98	69	654
266	1	2	1018	dbj|D14711|STAH	Staphylococcus aureus HSP10 and HSP60 genes	98	743	1017
282	1	1	525	gb|f.72488|	hemB=porphobilinogen synthase (Staphylococcus aureus, SA1959, Genomic, 1087 nt)	100	110	525
282	2	516	1502	gb|S72488|	hemB=porphobilinogen synthase (Staphylococcus aureus, SA1959, Genomic, 1087 nt)	100	952	987
284	1	3	170	gb|M63176|	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	98	84	168
284	2	282	1034	gb|M63176|	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	100	712	753
284	3	1028	2026	gb|M63176|	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	99	979	999
284	4	1990	2202	gb|M63176|	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	98	187	213
289	3	1536	1991	gb|M32470|	S.aureus Sau3AI-restriction-enzyme and Sau3AI-modification-enzyme genes, complete cds	99	338	456
303	1	2	868	gb|L01055|	Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hglC) genes, complete cds	99	867	867
303	2	1409	2383	gb|L01055|	Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hglC) genes, complete cds	100	975	975
303	3	2367	3161	gb|L01055|	Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hglC) genes, complete cds	99	793	795
305	1	2707	1355	dbj|D17366|STAA	Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	99	1343	1353
311	1	2628	1315	gb|L42945|	Staphylococcus aureus lytS and lytR genes, complete cds	98	1314	1314
312	6	7019	7870	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	74	351	852
323	1	1998	1003	gb|U31175|	Staphylococcus aureus D-specific D-2-hydroxyacid dehydrogenase (ddh) gene, complete cds	98	996	996
326	1	1	237	emb|Y00356|SASP	Staphylococcus aureus V8 serine protease gene	100	108	237
338	1	687	388	emb|X64389|SALE	S.aureus leuF-PB3 gene for F component of leucocidin R	98	259	300
338	2	1828	1088	emb|X64389|SALE	S.aureus leuF-PB3 gene for F component of leucocidin R	97	137	741

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name ( )	percent ident	HSP nt length	ORF nt length
342	2	579	1754	|gb|U06462|	Staphylococcus aureus SA4 FtsZ (ftsZ) gene, complete cds	100	1176	1176
344	2	517	1248	|emb|V01281|SANU	S.aureus mRNA for nuclease	98	732	732
349	1	457	230	|gb|M20393|	S.aureus bacteriophage phi-11 attachment site (attB)	96	172	228
353	1	1016	516	|gb|M83994|	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	100	187	501
353	2	1582	1046	|gb|M83994|	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	99	537	537
356	1	3	674	|gb|U20503|	Staphylococcus aureus MHC class II analog gene, complete cds	75	671	672
361	1	1	903	|gb|L19298|	Staphylococcus aures phosphatidylinositol-specific phospholipase C (plc) gene, complete cds	98	747	903
361	2	1103	1507	|gb|L19298|	Staphylococcus aures phosphatidylinositol-specific phospholipase C (plc) gene, complete cds	97	68	405
373	1	3	1148	|emb|X62288|SAPE	S.aureus DNA for penicillin-binding protein 2	99	1146	1146
389	3	1904	1248	|emb|X62282|SATS	S.aureus target site DNA for IS431 insertion	97	349	657
400	1	1	540	|emb|X61716|SAHL	S.aureus hlb gene encoding sphingomyelinase	99	389	540
400	2	1693	1187	|emb|X13404|SAHL	Staphylococcus aureus hlb gene for beta-hemolysin	99	178	507
408	1	1810	1049	|gb|S76213|	asp23=alkaline shock protein 23 (methicillin resistant) [Staphylococcus aureus, 912, Genomic, 1360 nt]	99	163	762
418	1	2	217	|gb|L41499|	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	100	216	216
418	2	854	639	|dbj|D17366|STAA	Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	100	188	216
421	2	1262	2509	|gb|L43098|	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	99	1248	1248
422	1	2	325	|gb|K02985|	S.aureus (strain RN450) transposon Tn554 insertion site	96	200	324
427	1	865	434	|dbj|D28879|STAP	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	100	432	432
427	2	1829	1122	|dbj|D28879|STAP	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	100	151	708
435	1	2	808	|dbj|D86240|D862	Staphylococcus aureus gene for unkown function and dlt operon dltA, dltB, dltC and dltD genes,complete cds	100	556	807
435	2	832	999	|dbj|D86240|D862	Staphylococcus aureus gene for unkown function and dlt operon dltA, dltB, dltC and dltD genes,complete cds	100	134	168
436	1	1341	685	|emb|X17688|SAFE	S.aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	97	657	657

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
436	2	2403	1657	emb|X17688|SAFE	S.aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	100	294	747
442	1	347	1300	emb|X72700|SAPV	S.aureus genes for S and F components of Panton-Valentine leucocidins	86	204	954
445	2	1906	2178	gb|L01055|	Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hglC) genes, complete cds	98	187	273
447	1	167	1078	gb|U19770|	Staphylococcus aureus pyrrolidone carboxyl peptidase (pcp) gene, complete cds	100	51*	912
447	2	1176	1784	gb|U19770|	Staphylococcus aureus pyrrolidone carboxyl peptidase (pcp) gene, complete cds	96	597	609
454	3	7309	4319	emb|Z18852|SACF	S.aureus gene for clumping factor	75	653	2991
472	4	7896	5479	gb|L25288|	Staphylococcus aureus gyrase-like protein alpha and beta subunit (grlA and grlB) genes, complete cds	99	2418	2418
472	5	8120	6792	gb|L25288|	Staphylococcus aureus gyrase-like protein alpha and beta subunit (grlA and grlB) genes, complete cds	99	1328	1329
475	2	566	889	emb|X52543|SAAG	S.aureus agrA, agrB and hld genes	100	76	324
481	4	1922	1560	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	250	363
481	5	1244	1534	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	224	291
487	2	1388	1188	gb|M83994|	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	98	72	201
489	1	2737	1370	gb|U21221|	Staphylococcus aureus hyaluronate lyase (hysA) gene, complete cds	99	1368	1368
503	2	1135	653	gb|M83994|	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	100	108	483
511	3	1613	2242	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	84	323	630
511	4	3122	2700	gb|S76213|	asp23=alkaline shock protein 23 (methicillin resistant) [Staphylococcus aureus, 912, Genomic, 1360 nt]	96	423	423
520	2	758	1297	emb|X72014|SAFI	S.aureus fib gene for fibrinogen-binding protein	99	540	540
520	3	1436	1801	emb|X72013|SAFI	S.aureus fib gene for fibrinogen-binding protein	99	221	366
526	1	2150	1092	dbj|D17366|STAA	Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	99	641	1059

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
528	2	58	963	gb|L19300|	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	99	260	906
528	3	1098	2870	gb|L19300|	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	99	866	1773
530	1	3	434	gb|U31979|	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroauinate synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	99	432	432
530	2	1211	2395	gb|U31979|	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroauinate synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	91	1185	1185
530	3	2409	2801	gb|U31979|	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroauinate synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	88	181	393
530	4	2690	3484	gb|L05004|	Staphylococcus aureus dehydroquinate synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	100	75	795
530	5	3482	4792	gb|L05004|	Staphylococcus aureus dehydroquinate synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	99	905	1311
530	6	4790	5380	gb|L05004|	Staphylococcus aureus dehydroquinate synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	100	196	591
539	1	3	338	emb|X76490|SAGL	S.aureus (bb270) glnA and glnR genes	99	336	336
539	2	336	527	emb|X76490|SAGL	S.aureus (bb270) glnA and glnR genes	100	189	192
554	1	727	365	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	100	54	363
554	2	2175	1252	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	918	924
554	3	1574	1374	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	96	122	201
584	2	1019	705	gb|U21221|	Staphylococcus aureus hyaluronate lyase (hysA) gene, complete cds	99	306	315
587	3	1475	4288	emb|Z18852|SACP	S.aureus gene for clumping factor	98	2588	2814
598	1	3881	1953	dbj|D28879|STAP	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	99	1873	1929

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
605	1	2	745	dbj|D86240|D862	Staphylococcus aureus gene for unkown function and dlt operon dltA, dltB, dltC and dltD genes,complete cds	98	338	744
609	1	1628	816	emb|X76490|SAGL	S.aureus (bb270) glnA and glnR genes	100	495	813
614	1	1280	642	gb|M32103|	Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA), partial cds	99	639	639
626	1	2508	1255	gb|M63176|	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	100	225,	1254
626	2	3315	2284	gb|M63176|	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	99	838	1032
629	1	1999	1001	emb|X17688|SAFE	S.aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	99	990	999
629	2	1407	1195	emb|X17688|SAFE	S.aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	98	194	213
631	2	5126	3228	emb|Z18852|SACF	S.aureus gene for clumping factor	82	489	1899
632	1	3	551	emb|Z30588|SAST	S.aureus (RN4220) genes for potential ABC transporter and potential membrane spanning protein	99	549	549
632	2	529	1323	emb|Z30588|SAST	S.aureus (RN4220) genes for potential ABC transporter and potential membrane spanning protein	99	795	795
651	1	1909	1070	gb|L19300|	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	99	478	840
657	2	1800	1105	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	84	456	696
662	1	908	456	emb|X13404|SAHL	Staphylococcus aureus hlb gene for beta-hemolysin	100	369	453
662	2	230	475	emb|X13404|SAHL	Staphylococcus aureus hlb gene for beta-hemolysin	100	246	246
662	3	746	1399	emb|X13404|SAHL	Staphylococcus aureus hlb gene for beta-hemolysin	99	653	654
682	1	956	480	gb|M63177|	S.aureus sigma factor (plaC) gene, complete cds	100	136	477
685	1	1182	592	gb|U65000|	Staphylococcus aureus type-I signal peptidase SpsA (spsA) gene, and type-I signal peptidase SpsB (spsB) gene, complete cds	98	534	591
685	2	1716	1153	gb|U65000|	Staphylococcus aureus type-I signal peptidase SpsA (spsA) gene, and type-I signal peptidase SpsB (spsB) gene, complete cds	96	564	564
697	1	3	527	gb|M63177|	S.aureus sigma factor (plaC) gene, complete cds	100	195	525
697	2	485	784	gb|M63177|	S.aureus sigma factor (plaC) gene, complete cds	97	280	300

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
710	1	15	503	dbj|D86240|D862	Staphylococcus aureus gene for unkown function and dlt operon dltA, dltB, dltC and dltD genes,complete cds	99	217	489
733	1	26	205	gb|M80252|	Staphylococcus aureus norA1199 gene (which mediates active efflux of fluoroquinolones), complete cds	97	140	180
741	1	1736	1197	dbj|D83951|STAL	Staphylococcus aureus DNA for LukM component, LukF-PV like component, complete cds	81	522	540
752	1	1	636	emb|Y00356|SASP	Staphylococcus aureus V8 serine protease gene	99	618	636
752	2	588	956	emb|Y00356|SASP	Staphylococcus aureus V8 serine protease gene	99	340	369
756	1	1308	709	emb|X01645|SATO	Staphylococcus aureus (Wood 46) gene for alpha-toxin	98	567	600
777	1	1582	950	emb|Z49245|SA42	S.aureus partial sod gene for superoxide dismutase	99	429	633
780	1	1111	557	gb|U20503|	Staphylococcus aureus MHC class-II analog gene, complete cds	86	550	555
784	1	73	687	gb|U63529|	Staphylococcus aureus novel antigen gene, complete cds	99	568	615
797	1	182	544	dbj|D14711|STAH	Staphylococcus aureus HSP10 and HSP60 genes	98	363	363
798	1	532	302	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	95	196	231
823	1	3	467	gb|S77055|	recF cluster: dnaA=replisome assembly protein...gyrB=DNA gyrase beta subunit [Staphylococcus aureus, YB886, Genomic, 5 genes, 3573 nt]	99	156	465
848	1	348	175	gb|L25288|	Staphylococcus aureus gyrase-like protein alpha and beta subunit (grlA and grlB) genes, complete cds	99	174	174
848	2	476	318	gb|L25288|	Staphylococcus aureus gyrase-like protein alpha and beta subunit (grlA and grlB) genes, complete cds	100	131	159
866	1	792	397	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	395	396
883	1	1	285	dbj|D90119|STAN	S. aureus norA gene	99	131	285
884	1	606	334	emb|X52543|SAAG	S.aureus agrA, agrB and hld genes	98	265	273
884	2	716	522	emb|X52543|SAAG	S.aureus agrA, agrB and hld genes	100	195	195
912	2	517	681	emb|Z30588|SAST	S.aureus (RN4220) genes for potential ABC transporter and potential membrane spanning protein	99	163	165
917	1	2	265	gb|M64724|	S.aureus tagatose 6-phosphate isomerase gene, complete cds	99	247	264
917	2	238	396	gb|M64724|	S.aureus tagatose 6-phosphate isomerase gene, complete cds	95	147	159
918	1	2426	1215	emt|X93205|SAPT	S.aureus ptsH and ptsI genes	99	1212	1212

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
967	1	1	411	dbj|D90119|STAN	S. aureus norA gene	97	395	411
991	1	672	337	emb|X52543|SAAG	S.aureus agrA, agrB and hld genes	99	336	336
1000	1	1117	845	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	78	190	273
1001	1	498	265	dbj|D86240|D862	Staphylococcus aureus gene for unkown function and dlt operon dltA, dltB, dltC and dltD genes,complete cds	99	234	234
1010	1	1	285	gb|U21221|	Staphylococcus aureus hyaluronate lyase (hysA) gene, complete cds	99	224	285
1046	1	656	330	emb|X72700|SAPV	S.aureus genes for S and F components of Panton-Valentine leucocidins	85	205	327
1060	1	480	286	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	99	180	195
1073	1	1176	589	gb|K02985|	S.aureus (strain RN450) transposon Tn554 insertion site	100	131	588
1079	1	3	230	dbj|D86240|D862	Staphylococcus aureus gene for unkown function and dlt operon dltA, dltB, dltC and dltD genes,complete cds	99	228	228
1079	2	218	484	dbj|D86240|D862	Staphylococcus aureus gene for unkown function and dlt operon dltA, dltB, dltC and dltD genes,complete cds	100	267	267
1079	3	460	645	dbj|D86240|D862	Staphylococcus aureus gene for unkown function and dlt operon dltA, dltB, dltC and dltD genes,complete cds	100	186	186
1092	1	289	146	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	98	124	144
1143	1	1	243	gb|M63177|	S.aureus sigma factor (plaC) gene, complete cds	99	243	243
1157	1	2	136	emb|Z48003|SADN	S.aureus gene for DNA polymerase III	97	127	135
1189	1	720	361	gb|S74031|	norA=NorA (ISP794) [Staphylococcus aureus, NCTC 8325, Insertion, 1820 nt]	99	360	360
1190	1	2	283	gb|M21854|	S.aureus agr gene encoding an accessory gene regulator protein, complete cds	100	282	282
1190	2	1127	888	emb|X52543|SAAG	S.aureus agrA, agrB and hld genes	100	240	240
1225	1	2	163	emb|X17679|SACO	Staphylococcus aureus coa gene for coagulase	97	124	162
1243	1	2	529	dbj|D86240|D862	Staphylococcus aureus gene for unkown function and dlt operon dltA, dltB, dltC and dltD genes,complete cds	99	495	528
1244	1	1	210	gb|S74031|	norA=NorA (ISP794) [Staphylococcus aureus, NCTC 8325, Insertion, 1820 nt]	100	210	210
1301	1	41	472	emb|X76490|SAGL	S.aureus (bb270) glnA and glnR genes	99	299	432

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
1315	1	18	326	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	98	277	309
1519	1	2	175	dbj|D28879|STAP	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	98	139	174
1663	1	1346	675	dbj|DA6240|D862	Staphylococcus aureus gene for unkown function and dlt operon dltA, dltB, dltC and dltD genes,complete cds	98	672	672
1797	1	644	324	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	321	321
1857	1	1	192	gb|M90536|	Staphylococcus aureus alpha-hemolysin gene, 3' end	98	192	192
1921	1	2	181	emb|X17688|SAFE	S.aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	100	180	180
1957	1	2	346	gb|U60589|	Staphylococcus aureus novel antigen gene, complete cds	99	345	345
1988	1	1	402	dbj|D86240|D862	Staphylococcus aureus gene for unkown function and dlt operon dltA, dltB, dltC and dltD genes,complete cds	100	402	402
2100	1	414	208	gb|M63177|	S.aureus sigma factor (plaC) gene, complete cds	99	207	207
2199	1	1	402	gb|U66664|	Staphylococcus aureus DNA fragment with class II promoter activity	99	131	402
2537	1	308	156	emb|X17688|SAFE	S.aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	99	153	153
2891	1	2	400	gb|L25426|	Staphylococcus aureus penicillin-binding protein 2 (pbp2) gene, complete cds	99	399	399
2950	1	778	398	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	100	358	381
2971	1	3	398	gb|U51132|	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (menE), and o-succinylbenzoic acid synthetase (menC) genes, complete cds	97	272	396
2978	1	618	328	gb|U31979|	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroquinate synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	98	250	291
2985	1	832	464	emb|X17679|SACO	Staphylococcus aureus coa gene for coagulase	98	347	369
3006	1	2170	1784	gb|U11779|	Staphylococcus aureus methicillin-resistant ATCC 33952 clone RRNV30 16S-23S rRNA spacer region	87	82	387
3008	1	474	238	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	88	178	237
3008	2	451	281	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	97	120	171

TABLE 1

S aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP length	ORF nt length
3011	1	793	398	emb|X62992|SAFN	S.aureus fnbB gene for fibronectin binding protein B	93	72	396
3019	1	2	235	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	97	234	234
3023	1	81	233	gb|U06451|	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	87	100	153
3029	1	90	287	gb|U51133|	Staphylococcus aureus phosphoenolpyruvate carboxykinase (pcka) gene, complete cds	100	135	198
3039	1	18	164	gb|U51133|	Staphylococcus aureus phosphoenolpyruvate carboxykinase (pcka) gene, complete cds	97	135	147
3039	2	70	327	gb|U51133|	Staphylococcus aureus phosphoenolpyruvate carboxykinase (pcka) gene, complete cds	77	183	258
3056	1	3	215	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	213	213
3059	1	1	261	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	98	234	261
3073	1	27	284	gb|U06451|	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	99	229	258
3074	1	2	397	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	96	250	396
3088	1	3	239	dbj|D86727|D867	Staphylococcus aureus DNA for DNA polymerase III, complete cds	95	215	237
3097	1	444	244	emb|Z48003|SADN	S.aureus gene for DNA polymerase III	97	160	201
3102	1	307	155	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	97	142	153
3121	1	568	398	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	100	88	171
3135	1	463	233	emb|X89233|SAMP	S.aureus DNA for rpoC gene	98	192	231
3133	1	2	175	emb|Z18852|SACF	S.aureus gene for clumping factor	96	154	174
3160	1	420	211	dbj|D10489|STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	89	197	210
3176	1	1	378	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	96	91	378
3192	1	420	211	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	98	72	210
3210	1	3	143	gb|U76714|	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	96	141	141

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
3232	3	2106	1282	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	71	257	825
3538	1	2	394	emb|X89233|SARP	S.aureus DNA for rpoC gene	99	356	393
3543	1	392	634	gb|L11530|	Staphylococcus aureus transfer RNA sequence with two rRNAs	99	102	243
3555	1	637	320	emb|Z18852|SACF	S.aureus gene for clumping factor	99	307	318
3559	1	3	182	emb|X17679|SACO	Staphylococcus aureus coa gene for coagulase	100	141	180
3559	2	95	313	emb|X17679|SACO	Staphylococcus aureus coa gene for coagulase	98	174	219
3563	1	278	141	gb|U35773|	Staphylococcus aureus prolipoprotein diacylglyceryl transferase (lgt) gene, complete cds	100	79	138
3563	2	527	363	gb|U35773|	Staphylococcus aureus prolipoprotein diacylglyceryl transferase (lgt) gene, complete cds	98	162	165
3566	1	3	422	emb|X16457|SAST	Staphylococcus aureus gene for staphylocoagulase	98	175	420
3588	1	2	262	gb|L43098|	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	99	253	261
3593	1	1	350	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	99	345	348
3600	1	758	381	emb|Z18852|SACF	S.aureus gene for clumping factor	72	346	378
3602	1	788	396	emb|Z18852|SACF	S.aureus gene for clumping factor	98	319	393
3656	1	1013	528	emb|Z18852|SACF	S.aureus gene for clumping factor	84	403	486
3682	1	3	236	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	231	234
3682	2	224	415	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	112	192
3693	1	758	423	emb|X62992|SAFN	S.aureus fnbB gene for fibronectin binding protein B	100	229	336
3702	1	593	354	gb|L11530|	Staphylococcus aureus transfer RNA sequence with two rRNAs	94	81	240
3725	1	924	463	emb|Z18852|SACF	S.aureus gene for clumping factor	71	367	462
3761	1	809	450	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	85	333	360
3767	1	1	402	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	98	387	402

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
3775	1	2	286	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	227	285
3786	1	456	229	dbj|D10489|STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	100	204	228
3786	2	542	366	dbj|D10489|STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	95	123	177
3798	1	3	251	emb|X17679|SACO	Staphylococcus aureus coa gene for coagulase	99	249	249
3813	1	793	398	gb|J04151|	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	98	396	396
3819	1	184	402	emb|X68425|SA23	S.aureus gene for 23S rRNA	99	161	219
3844	1	932	468	gb|U48826|	Staphylococcus aureus elastin binding protein (ebpS) gene, complete cds	87	204	465
3845	1	1	381	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	94	356	381
3856	1	798	400	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	76	192	399
3859	1	1049	573	emb|Z18852|SACP	S.aureus gene for clumping factor	85	347	477
3871	1	650	327	gb|M76714|	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	100	299	324
3876	1	2	253	dbj|D10489|STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	100	217	252
3877	1	572	288	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	97	209	285
3878	1	1	237	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate docarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	96	155	237
3888	1	3	173	emb|X16457|SAST	Staphylococcus aureus gene for staphylocoagulase	98	171	171
3893	1	1	183	emb|X89233|SARP	S.aureus DNA for rpoC gene	100	170	183
3893	2	181	357	emb|X89233|SARP	S.aureus DNA for rpoC gene	98	79	177
3894	1	3	485	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	450	483
3895	1	836	420	gb|J04151|	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	99	411	417
3905	1	48	239	gb|L05004|	Staphylococcus aureus dehydroquinate synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	100	159	192
3905	2	188	400	gb|L05004|	Staphylococcus aureus dehydroquinate synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	97	88	213

TABLE 1

EP 0 786 519 A2

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
3910	1	3	359	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	99	278	357
3915	1	1	330	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	75	175	330
3964	1	691	347	emb|Z48003|SADN	S.aureus gene for DNA polymerase III	100	295	345
4007	1	199	390	emb|X16457|SAST	Staphylococcus aureus gene for staphylocoagulase	98	163	192
4036	1	3	371	dbj|D10489|STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	99	339	369
4046	1	692	348	emb|Z18852|SACF	S.aureus gene for clumping factor	87	221	345
4060	1	1	375	emb|Z18852|SACF	S.aureus gene for clumping factor	96	271	375
4061	1	860	432	emb|Z48003|SADN	S.aureus gene for DNA polymerase III	99	429	429
4062	1	606	304	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	75	198	303
4085	1	58	402	gb|U11786|	Staphylococcus aureus methicillin-resistant ATCC 33952 clone RRUV42 16S-23S rRNA spacer region	98	127	345
4088	1	2	301	gb|L43098|	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	99	227	300
4093	1	2	277	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	99	276	276
4097	1	1	402	emb|Z18852|SACF	S.aureus gene for clumping factor	74	307	402
4116	1	22	402	gb|L05004|	Staphylococcus aureus dehydroquinate synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	98	157	381
4125	1	240	401	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	100	86	162
4149	1	35	247	gb|J04151|	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	99	200	213
4151	1	629	366	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	87	150	264
4154	1	754	398	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	297	357
4179	1	1	294	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	98	240	294

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match Accession	match gene name	percent ident	HSP nt length	ORF nt length
4203	1	1	255	emb|X89233|SARP	S.aureus DNA for rpoC gene	99	239	255
4206	1	1	303	emb|Z18852|SACF	S.aureus gene for clumping factor	100	236	303
4206	2	195	344	emb|Z18852|SACF	S.aureus gene for clumping factor	95	65	150
4208	1	108	314	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	89	76	207
4216	1	656	330	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	98	326	327
4226	1	594	298	gb|L11530|	Staphylococcus aureus transfer RNA sequence with two rRNAs	97	132	297
4260	1	216	383	gb|U11784|	Staphylococcus aureus methicillin-resistant ATCC 33952 clone RRNV40 16S-23S rRNA spacer region	83	141	168
4272	1	355	179	emb|Z48003|SADN	S.aureus gene for DNA polymerase III	100	164	177
4276	1	4	177	emb|X16457|SAST	Staphylococcus aureus gene for staphylocoagulase	99	150	174
4277	1	1	270	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	265	270
4282	1	691	377	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	98	282	315
4291	1	379	191	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	183	189
4295	1	3	329	emb|X16457|SAST	Staphylococcus aureus gene for staphylocoagulase	94	144	327
4313	1	435	280	gb|L11530|	Staphylococcus aureus transfer RNA sequence with two rRNAs	100	94	156
4315	1	3	185	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	100	158	183
4315	2	101	310	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	98	75	210
4327	1	1	294	gb|L43098|	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	98	294	294
4360	1	603	319	gb|U02910|	Staphylococcus aureus ATCC 25923 16S rRNA gene, partial sequence	100	116	285
4364	1	3	146	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	95	140	144
4388	1	167	310	emb|X62992|SAFN	S.aureus fnbB gene for fibronectin binding protein B	73	119	144

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
4401	1	2	313	emb|X62992|SAFN	S.aureus fnbB gene for fibronectin binding protein B	97	243	312
4421	1	36	281	dbj|D12572|STA2	Staphylococcus aureus rrnA gene for 23S ribosomal RNA	100	112	246
4426	1	3	293	emb|Z18852|SACF	S.aureus gene for clumping factor	85	185	291
4428	1	493	248	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	139	246
4462	1	2	271	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	270	270
4466	1	1	240	emb|Z18852|SACF	S.aureus gene for clumping factor	99	231	240
4469	1	1	312	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	99	265	312
4485	1	3	263	gb|L43098|	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	98	259	261
4492	1	74	400	gb|M86227|	Staphylococcus aureus DNA gyrase B subunit (gyrB) RecF homologue (recF) and DNA gyrase A subunit (gyrA) gene, complete cds	85	104	327
4497	1	515	269	emb|Z18852|SACF	S.aureus gene for clumping factor	99	213	267
4529	1	2	172	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	151	171
4547	1	1	300	emb|X62992|SAFN	S.aureus fnbB gene for fibronectin binding protein B	100	157	300
4554	1	318	160	emb|Z18852|SACF	S.aureus gene for clumping factor	84	126	159
4565	1	9	227	emb|Z18852|SACF	S.aureus gene for clumping factor	84	213	219
4569	1	79	222	emb|Z18852|SACF	S.aureus gene for clumping factor	98	127	144
4608	1	22	216	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	92	168	195
4614	1	464	234	emb|Z18852|SACF	S.aureus gene for clumping factor	86	169	231
4623	1	105	302	gb|J04151|	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	99	152	198
4632	1	18	206	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	98	183	189
4646	1	1	222	emb|Z18852|SACF	S.aureus gene for clumping factor	84	100	222
4687	1	2	166	gb|J04151|	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	98	156	165

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
4695	1	313	158	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	75	155	156
4703	1	1	153	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	98	103	153

TABLE 1

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
20	6	5089	4679	gi|511839	ORF1 [Staphylococcus bacteriophage phi 11]	100	100	411
149	3	2032	1577	pir|B49703|B497	int gene activator RinA - bacteriophage phi 11	100	100	456
149	5	2109	1912	gi|166161	Bacteriophage phi-11 int gene activator [Staphylococcus acteriophage phi 11]	100	100	198
349	2	558	409	gi|166159	integrase (int) [Staphylococcus bacteriophage phi 11]	100	100	150
398	1	1372	707	gi|166159	integrase (int) [Staphylococcus bacteriophage phi 11]	100	99	666
398	2	783	1001	gi|455128	excisionase (xis) [Staphylococcus bacteriophage phi 11]	100	100	219
502	4	1914	1744	gi|1204912	H. influenzae predicted coding region HI0660 [Haemophilus influenzae]	100	71	171
849	1	2	262	gi|1373002	polyprotein [Bean common mosaic virus]	100	46	261
1349	1	277	140	gi|143359	protein synthesis initiation factor 2 (infB) [Bacillus subtilis] gi|49319 IF2 gene product [Bacillus subtilis]	100	82	138
2880	1	21	308	gi|862933	protein kinase C inhibitor-I [Homo sapiens]	100	98	288
3085	1	428	216	gi|1354211	PET112-like protein [Bacillus subtilis]	100	100	213
4168	2	571	398	gi|1354211	PET112-like protein [Bacillus subtilis]	100	100	174
331	1	2	247	gi|426473	nusG gene product [Staphylococcus carnosus]	98	95	246
207	2	1272	1463	gi|460259	enolase [Bacillus subtilis]	97	90	192
331	2	395	850	gi|581638	L11 protein [Staphylococcus carnosus]	97	93	456
366	1	39	215	gi|166161	Bacteriophage phi-11 int gene activator [Staphylococcus acteriophage phi 11]	97	95	177
680	3	718	936	gi|426473	nusG gene product [Staphylococcus carnosus]	97	97	219
1578	1	284	144	gi|1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	97	79	141
157	1	321	518	gi|1022726	unknown [Staphylococcus haemolyticus]	96	88	198
205	33	16470	16147	gi|1165302	S10 [Bacillus subtilis]	96	91	324
3919	1	48	401	gi|971784	Clp-like ATP-dependent protease binding subunit [Bos taurus]	96	81	354
4133	1	830	417	gi|1022726	unknown [Staphylococcus haemolyticus]	96	84	414
4168	1	708	355	gi|1354211	PET112-like protein [Bacillus subtilis]	96	95	354
4207	1	312	157	gi|602031	similar to trimethylamine DH [Mycoplasma capricolum] pir|S49950|S49950 probable trimethylamine dehydrogenase (EC .5.99.7) - Mycoplasma capricolum (SGC3) (fragment)	96	86	156

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4227	2	152	331	gi|871784	Clp-like ATP-dependent protease binding subunit (Bos taurus)	96	81	180
4416	1	570	286	gi|1022726	unknown (Staphylococcus haemolyticus)	96	84	285
22	1	858	430	gi|511070	UreG (Staphylococcus xylosus)	95	88	429
22	7	4362	4036	gi|581787	urease gamma subunit (Staphylococcus xylosus)	95	79	327
82	6	8794	9114	pir|JG0008|JG00	ribosomal protein S7 - Bacillus stearothermophilus	95	83	321
154	9	9280	7838	gi|1354211	PET112-like protein (Bacillus subtilis)	95	92	1443
186	3	2798	2055	gi|1514656	serine O-acetyltransferase (Staphylococcus xylosus)	95	87	744
205	5	4406	4014	gi|142462	ribosomal protein S11 (Bacillus subtilis)	95	85	393
205	7	5017	4793	gi|142459	initiation factor 1 (Bacillus subtilis)	95	84	225
205	21	11365	10991	gi|1044974	ribosomal protein L14 (Bacillus subtilis)	95	93	375
259	5	7288	6644	sp|P47995|YSEA_	HYPOTHETICAL PROTEIN IN SECA 5'REGION (ORF1) (FRAGMENT).	95	85	645
302	3	795	1097	gi|40186	homologous to E.coli ribosomal protein L27 (Bacillus subtilis) i|143592 L27 ribosomal protein (Bacillus subtilis) ir|C21895|C21895 ribosomal protein L27 - Bacillus subtilis p|P05657|RL27_BACSU 50S RIBOSOMAL PROTEIN L27 (BL30) (BL24). i|40175 L24 gene prod	95	89	303
310	1	579	1523	gi|1177684	chorismate mutase (Staphylococcus xylosus)	95	92	945
414	1	2	163	pir|C48396|C483	ribosomal protein L34 - Bacillus stearothermophilus	95	90	162
4185	2	125	277	gi|1276841	glutamate synthase (GOGAT) (Porphyra purpurea)	95	86	153
22	2	1028	723	gi|511069	UreF (Staphylococcus xylosus)	94	91	306
22	5	5046	3310	gi|410516	urease alpha subunit (Staphylococcus xylosus)	94	85	1737
60	4	815	1372	gi|666116	glucose kinase (Staphylococcus xylosus)	94	87	558
205	18	10012	9536	gi|1044978	ribosomal protein S8 (Bacillus subtilis)	94	78	477
326	4	3378	2542	gi|557492	dihydroxynapthoic acid (DHNA) synthetase (Bacillus subtilis) gi|143186 dihydroxynapthoic acid (DHNA) synthetase (Bacillus ubtilis)	94	85	837
414	3	737	955	gi|467386	thiophen and furan oxidation (Bacillus subtilis)	94	77	219
426	3	2260	1823	gi|1263908	putative (Staphylococcus epidermidis)	94	87	438
534	1	2	355	gi|633650	enzyme II(mannitol) (Staphylococcus carnosus)	94	84	354
1017	1	2	229	gi|149435	putative (Lactococcus lactis)	94	73	228
3098	1	330	184	gi|413952	ipa-28d gene product (Bacillus subtilis)	94	50	147

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
3232	1	630	316	gi|1022725	unknown (Staphylococcus haemolyticus)	94	84	315
42	5	2089	2259	pir|B48396|B483	ribosomal protein L33 - Bacillus stearothermophilus	93	81	171
101	2	1745	1383	gi|155345	arsenic efflux pump protein (Plasmid pSX267)	93	82	363
205	24	12227	11865	sp|P14577|RL16_	50S RIBOSOMAL PROTEIN L16.	93	83	363
259	4	8291	5673	gi|499335	secA protein (Staphylococcus carnosus)	93	85	2619
275	1	2226	1114	gi|633650	enzyme II(mannitol) (Staphylococcus carnosus)	93	86	1113
444	6	6207	5773	gi|1022726	unknown (Staphylococcus haemolyticus)	93	81	435
491	1	152	622	gi|46912	ribosomal protein L13 (Staphylococcus carnosus)	93	88	471
607	6	1674	2033	gi|1022726	unknown (Staphylococcus haemolyticus)	93	83	360
653	1	973	488	gi|580890	translation initiation factor IF3 (AA 1-172) (Bacillus tearothermophilus)	93	77	486
1864	1	3	194	gi|306553	ribosmal protein small subunit (Homo sapiens)	93	93	192
2997	1	28	300	gi|143390	carbamyl phosphate synthetase (Bacillus subtilis)	93	82	273
3232	2	907	596	gi|1022725	unknown (Staphylococcus haemolyticus)	93	84	312
3761	2	794	621	gi|1022725	unknown (Staphylococcus haemolyticus)	93	88	174
16	1	3	374	gi|142781	putative cytoplasmic protein; putative (Bacillus subtilis) sp|P37954|UVRB_BACSU EXCINUCLEASE ABC SUBUNIT B (DINA PROTEIN) FRAGMENT).	92	83	372
31	7	5915	6124	gi|1136430	KIAA0185 protein (Homo sapiens)	92	46	210
56	19	26483	27391	gi|467401	unknown (Bacillus subtilis)	92	80	909
69	6	5882	6130	gi|530200	trophoblastin (Ovis aries)	92	53	249
145	3	2568	2038	gi|1022725	unknown (Staphylococcus haemolyticus)	92	80	531
171	3	2760	2362	gi|517475	D-amino acid transaminase (Staphylococcus haemolyticus)	92	86	399
205	12	7495	6962	gi|49189	secY gene product (Staphylococcus carnosus)	92	85	534
205	19	10812	10255	gi|1044976	ribosomal protein L5 (Bacillus subtilis)	92	82	558
219	1	710	357	gp|1303812	YqeV (Bacillus subtilis)	92	88	354
344	3	1575	1805	gi|1405474	CspC protein (Bacillus cereus)	92	85	231
699	1	20	361	gi|413999	ipa-75d gene product (Bacillus subtilis)	92	81	342
1343	1	2	160	pir|A45434|A454	ribosomal protein L19 - Bacillus stearothermophilus	92	84	159

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
1958	1	524	264	gi|407908	EIIscr (Staphylococcus xylosus)	92	80	261
3578	2	718	386	gi|1339950	large subunit of NADH-dependent glutamate synthase (Plectonema boryanum)	92	78	333
3585	1	644	324	gi|1339950	large subunit of NADH-dependent glutamate synthase (Plectonema boryanum)	92	81	321
3640	1	4	402	gi|1022726	unknown (Staphylococcus haemolyticus)	92	81	399
4362	1	14	178	gi|450688	hsdM gene of EcoprrI gene product (Escherichia coli) pir|S38437|S38437 hsdM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	92	78	165
4446	1	358	182	gi|1022725	unknown (Staphylococcus haemolyticus)	92	82	177
4549	1	462	232	gi|1022726	unknown (Staphylococcus haemolyticus)	92	80	231
4626	1	3	224	gi|1022725	unknown (Staphylococcus haemolyticus)	92	84	222
2	4	3980	4531	gi|535349	CodW (Bacillus subtilis)	91	74	552
28	1	2	1126	gi|1001376	hypothetical protein (Synechocystis sp.)	91	78	1125
60	5	1354	1701	gi|122604	orf2 downstream of glucose kinase (Staphylococcus xylosus)	91	80	348
101	1	1989	1036	gi|150728	arsenic efflux pump protein (Plasmid pI258)	91	80	954
187	2	412	1194	gi|142559	ATP synthase alpha subunit (Bacillus megaterium)	91	79	783
205	22	11579	11298	gi|40149	S17 protein (AA 1-87) (Bacillus subtilis)	91	83	282
206	7	8184	10262	gi|1072418	glcA gene product (Staphylococcus carnosus)	91	83	2079
306	2	3885	2326	gi|143012	GMP synthetase (Bacillus subtilis)	91	78	1560
306	3	5319	3826	gi|467399	IMP dehydrogenase (Bacillus subtilis)	91	79	1494
310	3	2194	3207	gi|1177685	ccpA gene product (Staphylococcus xylosus)	91	81	1014
343	4	2974	3150	gi|949974	sucrose repressor (Staphylococcus xylosus)	91	82	177
480	3	1606	3042	gi|433991	ATP synthase subunit beta (Bacillus subtilis)	91	85	1437
536	3	2026	1280	gi|143366	adenylosuccinate lyase (PUR-B) (Bacillus subtilis) pir|C29326|WZBSDS adenylosuccinate lyase (EC 4.3.2.2) - Bacillus ubtilis	91	79	747
552	1	1064	615	gi|297874	fructose-bisphosphate aldolase (Staphylococcus carnosus) pir|A49943|A49943 fructose-bisphosphate aldolase (EC 4.1.2.13) - taphylococcus carnosus (strain TM300)	91	79	450
637	1	1	1536	gi|143597	CTP synthetase (Bacillus subtilis)	91	79	1536
859	1	21	359	gi|385178	unknown (Bacillus subtilis)	91	66	339

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
1327	1	339	530	gi|496558	orfX [Bacillus subtilis]	91	71	192
2515	1	466	275	gi|511070	UreG [Staphylococcus xylosus]	91	85	192
2594	1	2	202	gi|146824	beta-cystathionase [Escherichia coli]	91	75	201
3764	1	847	425	gi|1022725	unknown [Staphylococcus haemolyticus]	91	78	423
4011	1	127	495	gi|1022726	unknown [Staphylococcus haemolyticus]	91	79	369
4227	1	1	177	gi|296464	ATPase [Lactococcus lactis]	91	66	177
42	3	815	1033	gi|520401	catalase [Haemophilus influenzae]	90	86	219
51	8	3717	4607	gi|580899	OppF gene product [Bacillus subtilis]	90	74	891
129	3	5317	4001	gi|1146206	glutamate dehydrogenase [Bacillus subtilis]	90	76	1317
164	17	16628	16933	sp|P05766|RS15_	30S RIBOSOMAL PROTEIN S15 (BS18).	90	74	306
171	5	2983	2819	gi|517475	D-amino acid transaminase [Staphylococcus haemolyticus]	90	78	165
205	4	4497	3550	gi|142463	RNA polymerase alpha-core-subunit [Bacillus subtilis]	90	76	948
205	6	4748	4410	gi|1044989	ribosomal protein S13 [Bacillus subtilis]	90	73	339
205	10	7165	6604	gi|49189	secY gene product [Staphylococcus carnosus]	90	81	762
205	11	6645	6472	gi|49189	secY gene product [Staphylococcus carnosus]	90	78	174
205	27	13692	13345	gi|786157	Ribosomal Protein S19 [Bacillus subtilis]	90	79	348
205	31	15858	15496	gi|1165303	L3 [Bacillus subtilis]	90	79	363
260	5	7023	5773	gi|1161380	IcaA [Staphylococcus epidermidis]	90	78	1251
299	6	3378	3947	gi|467440	'phosphoribosylpyrophosphate synthetase [Bacillus subtilis] gi|40218 PRPP synthetase (AA 1-317) [Bacillus subtilis]	90	78	570
320	2	1025	1717	gi|312443	carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus aldolyticus]	90	75	693
330	4	1581	1769	gi|986963	beta-tubulin [Sporidiobolus pararoseus]	90	80	189
369	1	954	523	pir|S34762|S347	L-serine dehydratase beta chain - Clostridium sp.	90	77	432
557	1	3	188	gi|1511589	M. jannaschii predicted coding region MJ1624 [Methanococcus jannaschii]	90	54	186
663	2	667	1200'	gi|143786	tryptophanyl-tRNA synthetase (EC 6.1.1.2) [Bacillus subtilis] pir|JT0481|YWBS tryptophan--tRNA ligase (EC 6.1.1.2) - Bacillus ubtilis	90	73	534
717	1	1	261	gi|143065	hubst [Bacillus stearothermophilus]	90	79	261
745	4	1059	865	gi|1205433	H. influenzae predicted coding region HI1190 [Haemophilus influenzae]	90	81	195

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
1007	1	386	565	gi|143366	adenylosuccinate lyase (PUR-B) (Bacillus subtilis) pir|C29326|WZBSDS adenylosuccinate lyase (EC 4.3.2.2) - Bacillus ubtilis	90	77	180
1054	1	579	331	gi|1033122	ORF_f729 (Escherichia coli)	90	50	249
1156	1	117	707	gi|1477776	ClpP (Bacillus subtilis)	90	80	591
1180	1	408	205	gi|1377831	unknown (Bacillus subtilis)	90	74	204
1253	1	1	462	gi|40046	phosphoglucose isomerase A (AA 1-449) (Bacillus stearothermophilus) ir|S15936|NUBSSA glucose-6-phosphate isomerase (EC 5.3.1.9) A - cillus stearothermophilus	90	75	462
2951	1	3	269	gi|144816	formyltetrahydrofolate synthetase (FTHFS) (ttg start codon) (EC .3.4.3) (Moorella thermoacetica)	90	76	267
3140	1	327	166	gi|1070014	protein-dependent (Bacillus subtilis)	90	52	162
4594	1	3	233	gi|871784	Clp-like ATP-dependent protease binding subunit (Bos taurus)	90	76	231
87	1	1028	1750	gi|467327	unknown (Bacillus subtilis)	89	75	723
112	1	2	505	gi|153741	ATP-binding protein (Streptococcus mutans)	89	77	504
118	1	120	398	gi|1303804	YqeQ (Bacillus subtilis)	89	75	279
128	4	3545	3757	gi|460257	triose phosphate isomerase (Bacillus subtilis)	89	84	213
164	12	11667	12755	gi|39954	IF2 (aa 1-741) (Bacillus stearothermophilus)	89	80	1089
205	13	7875	7405	gi|216338	ORF for L15 ribosomal protein (Bacillus subtilis)	89	76	471
205	32	16152	15823	gi|1165303	L3 (Bacillus subtilis)	89	80	330
270	3	2407	2207	pir|C41902|C419	arsenate reductase (EC 1.-.-.-) - Staphylococcus xylosus plasmid pSX267	89	81	201
395	2	157	672	gi|520574	glutamate racemase (Staphylococcus haemolyticus)	89	80	516
494	1	3	839	gi|396259	protease (Staphylococcus epidermidis)	89	77	837
510	1	1	444	gi|40046	phosphoglucose isomerase A (AA 1-449) (Bacillus stearothermophilus) ir|S15936|NUBSSA glucose-6-phosphate isomerase (EC 5.3.1.9) A - cillus stearothermophilus	89	74	444
615	1	2124	1210	gi|1303812	YqeV (Bacillus subtilis)	89	74	915
841	1	18	341	gi|1165303	L3 (Bacillus subtilis)	89	80	324
1111	1	352	813	gi|47146	thermonuclease (Staphylococcus intermedius)	89	70	462
1875	1	2	256	gi|1205108	ATP-dependent protease binding subunit (Haemophilus influenzae)	89	82	255
2963	1	11	367	gi|467458	cell division protein (Bacillus subtilis)	89	83	357

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
3020	1	90	362	gi|1239988	hypothetical protein (Bacillus subtilis)	89	66	273
3565	1	2	400	gi|1256635	dihydroxy-acid dehydratase (Bacillus subtilis)	89	75	399
3586	1	105	314	gi|580832	ATP synthase subunit gamma (Bacillus subtilis)	89	82	210
3629	1	794	399	gi|1009366	Respiratory nitrate reductase (Bacillus subtilis)	89	78	396
3688	1	2	400	gi|1146206	glutamate dehydrogenase (Bacillus subtilis)	89	75	399
3699	1	794	399	gi|1339950	large subunit of NADH-dependent glutamate synthase (Plectonema boryanum)	89	75	396
4016	1	428	216	gi|1009366	Respiratory nitrate reductase (Bacillus subtilis)	89	71	213
4177	1	471	301	gi|149426	putative (Lactococcus lactis)	89	76	171
4436	1	601	302	gi|1022725	unknown (Staphylococcus haemolyticus)	89	80	300
4635	1	320	162	gi|1022725	unknown (Staphylococcus haemolyticus)	89	73	159
2	2	1330	2676	gi|520754	putative (Bacillus subtilis)	88	76	1347
42	2	468	848	sp|P42321|CATA_	CATALASE (EC 1.11.1.6).	88	76	381
53	5	6389	4722	gi|474177	alpha-D-1,4-glucosidase (Staphylococcus xylosus)	88	80	1668
56	16	18018	18617	gi|467411	recombination protein (Bacillus subtilis)	88	77	600
60	3	376	843	gi|666116	glucose kinase (Staphylococcus xylosus)	88	77	468
70	2	1583	1245	gi|44095	replication initiator protein (Listeria monocytogenes)	88	74	339
82	8	11514	12719	pir|A60663|A606	translation elongation factor Tu - Bacillus subtilis	88	79	1206
103	7	4179	4391	gi|167181	serine/threonine kinase receptor (Brassica napus)	88	77	213
114	8	7732	8232	gi|1022726	unknown (Staphylococcus haemolyticus)	88	72	501
118	2	308	2011	gi|1303804	YqeQ (Bacillus subtilis)	88	77	1704
141	3	657	1136	gi|1405446	transketolase (Bacillus subtilis)	88	72	480
148	7	5871	6116	gi|1118002	dihydropteroate synthase (Staphylococcus haemolyticus)	88	78	246
165	3	1428	2231	gi|40053	phenylalanyl-tRNA synthetase alpha subunit (Bacillus subtilis) | ir|S11730|YFBSA phenylalanine--tRNA ligase (EC 6.1.1.20) alpha ain - | Bacillus subtilis	88	80	804
205	28	15027	14185	gi|1165306	L2 (Bacillus subtilis)	88	82	843
225	1	1569	898	gi|1303840	YqfS (Bacillus subtilis)	88	78	672
235	1	2	1975	gi|452309	valyl-tRNA synthetase (Bacillus subtilis)	88	76	1974

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
339	3	2060	1566	gi|1118002	dihydropteroate synthase (Staphylococcus haemolyticus)	88	73	495
443	4	4325	2928	gi|558559	pyrimidine nucleoside phosphorylase (Bacillus subtilis)	88	73	1398
532	1	3	419	gi|143797	valyl-tRNA synthetase (Bacillus stearothermophilus) sp|P11931|SYV_BACST VALYL-TRNA SYNTHETASE (EC 6.1.1.9) VALINE--TRNA LIGASE) (VALRS).	88	78	417
534	3	2504	2968	gi|153049	mannitol-specific enzyme-III (Staphylococcus carnosus) pir|JQ0088|JQ0088 phosphotransferase system enzyme II (EC .7.1.69), mannitol-specific, factor III - Staphylococcus carnosus sp|P17876|PTMA_STACA PTS SYSTEM, MANNITOL-SPECIFIC IIA COMPONENT EIIA-MTL) (	88	82	465
705	2	584	399	gi|710018	nitrite reductase (nirB) (Bacillus subtilis)	88	70	186
1000	2	1824	1309	gi|1022726	unknown (Staphylococcus haemolyticus)	88	78	516
1299	1	587	324	gi|401786	phosphomannomutase (Mycoplasma pirum)	88	55	264
1341	2	170	400	gi|39963	ribosomal protein L20 (AA 1-119) (Bacillus stearothermophilus) ir|S05348|R5BS20 ribosomal protein L20 - Bacillus earothermophilus	88	83	231
1386	1	41	214	pir|B47154|B471	signal recognition particle 54K chain homolog Ffh - Bacillus subtilis	88	71	174
1386	2	183	533	pir|B47154|B471	signal recognition particle 54K chain homolog Ffh - Bacillus subtilis	88	73	351
2949	1	704	399	gi|535350	CodX (Bacillus subtilis)	88	73	306
2984	1	5	169	gi|218277	O-acetylserine(thiol) lyase (Spinacia oleracea)	88	70	165
3035	1	1	138	gi|493083	dihydroxyacetone kinase (Citrobacter freundii)	88	67	138
3089	1	3	152	gi|606055	ORF_f746 (Escherichia coli)	88	88	150
3917	1	817	410	gi|143378	pyruvate decarboxylase (E-1) beta subunit (Bacillus subtilis) gi|1377836 pyruvate decarboxylase E-1 beta subunit (Bacillus ubtilis)	88	77	408
4199	1	680	342	gi|1405454	aconitase (Bacillus subtilis)	88	82	339
4201	1	734	369	gi|515938	glutamate synthase (ferredoxin) (Synechocystis sp.) pir|S46957|S46957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - ynechocystis sp.	88	84	366
4274	1	1	336	gi|515938	glutamate synthase (ferredoxin) (Synechocystis sp.) pir|S46957|S46957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - ynechocystis sp.	88	84	336
4308	1	794	399	gi|1146206	glutamate dehydrogenase (Bacillus subtilis)	88	71	396
2	5	4570	6000	gi|535350	CodX (Bacillus subtilis)	87	70	1431
52	8	6781	6482	gi|1064791	function umknown (Bacillus subtilis)	87	66	300

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
73	3	1584	2480	gi|142992	glycerol kinase (glpK) (EC 2.7.1.30) [Bacillus subtilis] pir|B45868|B45868 glycerol kinase (EC 2.7.1.30) - Bacillus subtilis sp|P18157|GLPK_BACSU GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL -PHOSPHOTRANSFERASE) (GLYCEROKINASE) (GK).	87	72	897
98	12	8813	9100	gi|467433	unknown [Bacillus subtilis]	87	62	288
124	4	4265	2988	gi|556886	serine hydroxymethyltransferase [Bacillus subtilis] pir|S49363|S49363 serine hydroxymethyltransferase - Bacillus ubtilis	87	77	1278
124	6	4457	4032	gi|556883	Unknown [Bacillus subtilis]	87	66	426
148	5	3741	4559	gi|467460	unknown [Bacillus subtilis]	87	70	819
164	13	12710	13810	gi|39954	IF2 (aa 1-741) [Bacillus stearothermophilus]	87	72	1101
177	2	1104	2126	gi|467385	unknown [Bacillus subtilis]	87	78	1023
199	1	1982	1158	gi|143527	iron-sulfur protein [Bacillus subtilis]	87	77	825
199	2	4717	2933	pir|A27763|A277	succinate dehydrogenase (EC 1.3.99.1) flavoprotein - Bacillus subtilis	87	80	1785
205	23	11782	11543	gi|1044972	ribosomal protein L29 [Bacillus subtilis]	87	78	240
205	25	13275	12607	gi|1165309	S3 [Bacillus subtilis]	87	75	669
222	1	2033	1107	gi|1177249	rec233 gene product [Bacillus subtilis]	87	70	927
236	3	1635	1333	gi|1146198	ferredoxin [Bacillus subtilis]	87	80	303
246	5	2585	2292	gi|467373	ribosomal protein S18 [Bacillus subtilis]	87	77	294
260	2	4189	3422	gi|1161382	IcaC [Staphylococcus epidermidis]	87	72	768
320	3	1696	2391	gi|312443	carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus aldolyticus]	87	80	696
380	4	1165	1383	gi|142570	ATP synthase c subunit [Bacillus firmus]	87	80	219
414	4	900	1073	gi|467386	thiophen and furan oxidation [Bacillus subtilis]	87	77	174
425	2	1003	794	gi|1046166	pilin repressor [Mycoplasma genitalium]	87	69	210
448	1	1255	722	gi|405134	acetate kinase [Bacillus subtilis]	87	75	534
480	1	1	711	gi|142559	ATP synthase alpha subunit [Bacillus megaterium]	87	79	711
481	1	2	352	sp|Q06797|RL1_B	50S RIBOSOMAL PROTEIN L1 (BL1).	87	72	351
677	2	359	955	gi|460911	fructose-bisphosphate aldolase [Bacillus subtilis]	87	78	597
677	3	934	1284	gi|460911	fructose-bisphosphate aldolase [Bacillus subtilis]	87	78	351
876	1	3	452	gi|1146247	asparaginyl-tRNA synthetase [Bacillus subtilis]	87	79	450

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
1376	1	426	214	gi|1065555	F46H6.4 gene product [Caenorhabditis elegans]	87	75	213
2206	1	3	374	gi|215098	excisionase [Bacteriophage 154a]	87	72	372
2938	1	3	290	gi|508979	GTP-binding protein [Bacillus subtilis]	87	69	288
3081	2	126	308	gi|467399	IMP dehydrogenase [Bacillus subtilis]	87	72	183
3535	1	3	401	gi|1405454	aconitase [Bacillus subtilis]	87	80	399
4238	1	547	275	gi|603769	HutU protein, urocanase [Bacillus subtilis]	87	73	273
4	8	10427	8736	gi|603769	HutU protein, urocanase [Bacillus subtilis]	86	72	1692
22	6	4190	3738	gi|410515	urease beta subunit [Staphylococcus xylosus]	86	73	453
54	2	2480	1572	gi|289287	UDP-glucose pyrophosphorylase [Bacillus subtilis]	86	70	909
124	3	2336	1713	gi|556887	uracil phosphoribosyltransferase [Bacillus subtilis] pir|S49364|S49364 uracil phosphoribosyltransferase - Bacillus ubtilis	86	74	624
148	3	1349	3448	gi|467458	cell division protein [Bacillus subtilis]	86	75	2100
148	4	3638	3859	gi|467460	unknown [Bacillus subtilis]	86	73	222
152	3	1340	2086	gi|1377835	pyruvate decarboxylase E-1 alpha subunit [Bacillus subtilis]	86	75	747
164	18	17347	19467	gi|1184680	polynucleotide phosphorylase [Bacillus subtilis]	86	72	2121
180	2	554	1159	gi|143467	ribosomal protein S4 [Bacillus subtilis]	86	80	606
205	3	2966	2592	gi|142464	ribosomal protein L17 [Bacillus subtilis]	86	77	375
205	26	13364	12990	gi|40107	ribosomal protein L22 [Bacillus stearothermophilus] ir|S10612|S10612 ribosomal protein L22 - Bacillus earothermophilus	86	75	375
246	7	3463	3140	gi|467375	ribosomal protein S6 [Bacillus subtilis]	86	70	324
299	3	1196	1540	gi|39656	spoVG gene product [Bacillus megaterium]	86	70	345
299	7	3884	4345	gi|467440	phosphoribosylpyrophosphate synthetase [Bacillus subtilis] gi|40218 PRPP synthetase (AA 1-317) [Bacillus subtilis]	86	78	462
304	5	2170	2523	gi|666983	putative ATP binding subunit [Bacillus subtilis]	86	65	354
310	2	1487	1678	gi|1177684	chorismate mutase [Staphylococcus xylosus]	86	71	192
337	5	2086	3405	gi|487434	isocitrate dehydrogenase [Bacillus subtilis]	86	78	1320
339	2	1489	1109	gi|1118003	dihydroneopterin aldolase [Staphylococcus haemolyticus]	86	77	381
358	2	2124	3440	gi|1146219	28.2% of identity to the Escherichia coli GTP-binding protein Era; putative [Bacillus subtilis]	86	73	1317

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
404	2	1015	2058	gi|1303817	YqfA (Bacillus subtilis)	86	70	1044
581	2	661	452	gi|40056	phoP gene product (Bacillus subtilis)	86	71	210
642	2	338	1075	gi|1176399	EpiF (Staphylococcus epidermidis)	86	72	738
770	1	622	347	gi|143328	phoP protein (put.); putative (Bacillus subtilis)	86	69	276
865	1	1777	890	gi|1146247	asparaginyl-tRNA synthetase (Bacillus subtilis)	86	74	888
868	2	963	1133	gi|1002911	transmembrane protein (Saccharomyces cerevisiae)	86	69	171
904	1	1	162	gi|1303912	YqhW (Bacillus subtilis)	86	72	162
989	1	35	433	gi|1303993	YqkL (Bacillus subtilis)	86	76	399
1212	1	296	150	gi|414014	ipa-90d gene product (Bacillus subtilis)	86	70	147
1323	1	2	148	gi|40041	pyruvate dehydrogenase (lipoamide) (Bacillus stearothermophilus) ir|S10798|DEBSPF pyruvate dehydrogenase (lipoamide) (EC 1.2.4.1) pha chain - Bacillus stearothermophilus	86	75	147
3085	2	540	310	gi|1354211	PET112-like protein (Bacillus subtilis)	86	86	231
3847	1	1	228	gi|296464	ATPase (Lactococcus lactis)	86	63	228
4487	1	476	240	gi|1022726	unknown (Staphylococcus haemolyticus)	86	73	237
4583	1	372	187	gi|1022725	unknown (Staphylococcus haemolyticus)	86	79	186
25	5	4287	5039	gi|1502421	3-ketoacyl-acyl carrier protein reductase (Bacillus subtilis)	85	64	753
56	21	30627	29395	gi|1408507	pyrimidine nucleoside transport protein (Bacillus subtilis)	85	69	1233
68	2	332	1192	gi|467376	unknown (Bacillus subtilis)	85	74	861
73	2	880	1707	gi|142992	glycerol kinase (glpK) (EC 2.7.1.30) (Bacillus subtilis) pir|B45868|B45868 glycerol kinase (EC 2.7.1.30) - Bacillus subtilis sp|P18157|GLPK_BACSU GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL -PHOSPHOTRANSFERASE) (GLYCEROKINASE) (GK).	85	72	828
106	4	1505	3490	gi|143766	(thrSv) (EC 6.1.1.3) (Bacillus subtilis)	85	74	1986
128	2	1153	2202	gi|311924	glyceraldehyde-3-phosphate dehydrogenase (Clostridium pasteurianum) pir|S34254|S34254 glyceraldehyde-3-phosphate dehydrogenase (EC .2.1.12) - Clostridium pasteurianum	85	75	1050
129	4	6466	5252	gi|1064807	ORTHININE AMINOTRANSFERASE (Bacillus subtilis)	85	73	1215
138	6	3475	5673	gi|1072419	glcB gene product (Staphylococcus carnosus)	85	74	2199
189	1	2	169	gi|467385	unknown (Bacillus subtilis)	85	65	168

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
205	15	8624	8106	gi|1044981	ribosomal protein S5 (Bacillus subtilis)	85	75	519
205	20	10928	10596	pir|A02819|R5BS	ribosomal protein L24 - Bacillus stearothermophilus	85	72	333
220	6	6490	6101	gi|48980	secA gene product (Bacillus subtilis)	85	66	390
231	4	4877	3159	gi|1002520	MutS (Bacillus subtilis)	85	70	1719
243	9	8013	8783	gi|414011	ipa-87r gene product (Bacillus subtilis)	85	72	771
249	2	5894	3186	gi|1405454	aconitase (Bacillus subtilis)	85	73	2709
302	1	140	475	gi|40173	homolog of E.coli ribosomal protein L21 (Bacillus subtilis) ir|S18439|S18439 Ribosomal protein L21 - Bacillus subtilis p|P26908|RL21_BACSU 50S RIBOSOMAL PROTEIN L21 (BL20).	85	72	336
333	1	5445	2968	gi|442360	ClpC adenosine triphosphatase (Bacillus subtilis)	85	69	2478
364	6	6082	8196	gi|871784	Clp-like ATP-dependent protease binding subunit (Bos taurus)	85	68	2115
448	2	1992	1339	gi|405134	acetate kinase (Bacillus subtilis)	85	68	654
747	1	1251	853	gi|1373157	orf-X; hypothetical protein; Method: conceptual translation supplied by author (Bacillus subtilis)	85	73	399
886	2	159	467	gi|541768	hemin permease (Yersinia enterocolitica)	85	55	309
1089	1	1208	606	pir|B47154|B471	signal recognition particle 54K chain homolog Ffh - Bacillus subtilis	85	71	603
1163	1	816	409	gi|304155	diaminopimelate decarboxylase (Bacillus methanolicus) sp|P41023|DCDA_BACMT DIAMINOPIMELATE DECARBOXYLASE (EC 4.1.1.20) DAP DECARBOXYLASE).	85	62	408
1924	1	487	251	gi|215098	excisionase (Bacteriophage 154a)	85	73	237
2932	1	776	390	gi|1041099	Pyruvate Kinase (Bacillus licheniformis)	85	71	387
3030	1	3	275	gi|42370	pyruvate formate-lyase (AA 1-760) (Escherichia coli) ir|S01788|S01788 formate C-acetyltransferase (EC 2.3.1.54) - cherichia coli	85	74	273
3111	1	595	299	gi|63568	limb deformity protein (Gallus gallus)	85	85	297
3778	1	630	316	gi|391840	beta-subunit of HDT (Pseudomonas fragi)	85	67	315
3835	1	1	387	gi|1204472	type I restriction enzyme ECOR124/3 I M protein (Haemophilus influenzae)	85	56	387
4042	1	3	386	gi|18178	formate acetyltransferase (Chlamydomonas reinhardtii) ir|S24997|S24997 formate C-acetyltransferase (EC 2.3.1.54) - lamydomonas reinhardtii	85	70	384
4053	1	35	340	gi|1204472	type I restriction enzyme ECOR124/3 I M protein (Haemophilus influenzae)	85	56	306
4108	1	2	181	gi|1072418	glcA gene product (Staphylococcus carnosus)	85	61	180
4300	1	575	330	gi|151932	fructose enzyme II (Rhodobacter capsulatus)	85	59	246

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name ( )	% sim	% ident	length (nt)
4392	1	627	355	gi|1022725	unknown (Staphylococcus haemolyticus)	85	74	273
4408	1	2	235	gi|871784	Clp-like ATP-dependent protease binding subunit (Bos taurus)	85	62	234
4430	1	578	291	gi|1009366	Respiratory nitrate reductase (Bacillus subtilis)	85	68	288
4555	1	2	253	gi|450688	hsdM gene of EcoprrI gene product (Escherichia coli) pir|S38437|S38437 hsdM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	85	52	252
4611	1	481	242	gi|1256635	dihydroxy-acid dehydratase (Bacillus subtilis)	85	65	240
4	10	10061	10591	gi|46982	fosB gene product (Staphylococcus epidermidis)	84	68	531
13	2	1348	1172	gi|142450	ahrC protein (Bacillus subtilis)	84	56	177
16	4	1803	4652	gi|1277198	DNA repair protein (Deinococcus radiodurans)	84	67	2850
22	3	1535	1128	gi|511069	UreF (Staphylococcus xylosus)	84	73	408
23	7	5055	5306	gi|603320	Yer082p (Saccharomyces cerevisiae)	84	61	252
53	11	11597	11145	gi|1303948	YqiW (Bacillus subtilis)	84	68	453
53	12	14059	12770	gi|142613	branched chain alpha-keto acid dehydrogenase E2 (Bacillus subtilis) gi|1303944 BfmBB (Bacillus subtilis)	84	71	1290
70	1	1332	982	gi|46647	ORF (repE) (Staphylococcus aureus)	84	68	351
73	4	2512	4311	gi|142993	glycerol-3-phosphate dehydrogenase (glpD) (EC 1.1.99.5) (Bacillus ubtilis)	84	74	1800
98	7	4324	6096	gi|467427	methionyl-tRNA synthetase (Bacillus subtilis)	84	66	1773
100	9	9501	8680	gi|1340128	ORF1 (Staphylococcus aureus)	84	78	822
117	3	1934	3208	gi|1237019	Srb (Bacillus subtilis)	84	68	1275
148	6	4720	5670	gi|467462	cysteine synthetase A (Bacillus subtilis)	84	69	951
152	4	2064	2456	gi|143377	pyruvate decarboxylase (E-1) alpha subunit (Bacillus subtilis) pir|B36718|DEBSPA pyruvate dehydrogenase (lipoamido) (EC 1.2.4.1) lpha chain - Bacillus subtilis	84	70	393
169	7	3634	3861	gi|1001342	hypothetical protein (Synechocystis sp.)	84	66	228
171	4	2992	2657	gi|517475	D-amino acid transaminase (Staphylococcus haemolyticus)	84	71	336
186	6	6941	6216	gi|467475	unknown (Bacillus subtilis)	84	70	726
205	9	6261	5692	gi|216340	ORF for adenylate kinase (Bacillus subtilis)	84	71	570
224	2	915	1391	gi|288269	beta-fructofuranosidase (Staphylococcus xylosus)	84	70	477

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name ( )	% sim	% identity	length (nt)
251	1	92	388	gi|1303790	YqeI [Bacillus subtilis]	84	65	297
282	3	1526	2836	gi|143040	glutamate-1-semialdehyde 2,1-aminotransferase [Bacillus subtilis] pir|D42728|D42728 glutamate-1-semialdehyde 2,1-aminomutase (EC .4.3.8) - Bacillus subtilis	84	75	1311
307	5	3138	2959	gi|1070014	protein-dependent [Bacillus subtilis]	84	62	180
320	4	2343	4229	gi|143390	carbamyl phosphate synthetase [Bacillus subtilis]	84	70	1887
372	1	3	296	gi|1022725	unknown [Staphylococcus haemolyticus]	84	70	294
413	2	2201	1341	gi|1256146	YbbQ [Bacillus subtilis]	84	65	861
439	1	3	392	gi|1046173	osmotically inducible protein [Mycoplasma genitalium]	84	53	390
461	3	1362	2270	gi|40211	threonine synthase (thrC) (AA 1-352) [Bacillus subtilis] ir|A25364|A25364 threonine synthase (EC 4.2.99.2) - Bacillus btilis	84	69	909
487	1	3	299	gi|1144531	integrin-like protein alpha Intlp [Candida albicans]	84	46	297
491	2	624	905	pir|S08564|R3BS	ribosomal protein S9 - Bacillus stearothermophilus	84	69	282
491	3	816	1033	pir|S08564|R3BS	ribosomal protein S9 - Bacillus stearothermophilus	84	77	198
548	1	3	341	gi|431231	uracil permease [Bacillus caldolyticus]	84	74	339
728	2	2701	1748	gi|912445	DNA polymerase [Bacillus caldotenax]	84	68	954
769	1	3	257	gi|1510953	cobalamin biosynthesis protein N [Methanococcus jannaschii]	84	38	255
954	1	308	156	gi|1405454	aconitase [Bacillus subtilis]	84	57	153
957	1	3	395	gi|143402	recombination protein (ttg start codon) [Bacillus subtilis] gi|1303923 RecN [Bacillus subtilis]	84	68	393
975	1	3	452	gi|885934	ClpB [Synechococcus sp.]	84	70	450
1585	1	3	257	gi|510140	ligoendopeptidase F [Lactococcus lactis]	84	56	255
2954	1	3	323	gi|603769	HutU protein, urocanase [Bacillus subtilis]	84	73	321
2996	1	650	348	gi|18178	formate acetyltransferase [Chlamydomonas reinhardtii] ir|S24997|S24997 formate C-acetyltransferase (EC 2.3.1.54) - lamydomonas reinhardtii	84	65	303
3766	1	737	375	gi|517205	67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogenes]	84	72	363
4022	1	2	169	gi|1146206	glutamate dehydrogenase [Bacillus subtilis]	84	54	168
4058	1	620	312	gi|151932	fructose enzyme II [Rhodobacter capsulatus]	84	71	309
4108	2	106	351	gi|1072418	glcA gene product [Staphylococcus carnosus]	84	77	246

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
4183	1	3	308	gi|603769	HutU protein, urocanase (Bacillus subtilis)	84	72	306
4726	1	55	234	gi|146208	glutamate synthase large subunit (EC 2.6.1.53) (Escherichia coli) pir|A29617|A29617 glutamate synthase (NADPH) (EC 1.4.1.13) large hain - Escherichia coli	84	73	180
22	4	2043	1576	gi|393297	urease accessory protein (Bacillus sp.)	83	64	468
53	13	14722	13745	gi|142612	branched chain alpha-keto acid dehydrogenase E1-beta (Bacillus ubtilis)	83	68	978
57	16	13357	12872	gi|143132	lactate dehydrogenase (AC 1.1.1.27) (Bacillus caldolyticus) pir|B29704|B29704 L-lactate dehydrogenase (EC 1.1.1.27) - Bacillus aldolyticus	83	66	486
66	3	3119	2274	gi|1303894	YqhM (Bacillus subtilis)	83	63	846
66	5	6118	4643	gi|1212730	YqhK (Bacillus subtilis)	83	68	1476
70	3	1864	1523	gi|44095	replication initiator protein (Listeria monocytogenes)	83	73	342
90	1	377	1429	gi|155571	alcohol dehydrogenase I (adhA) (EC 1.1.1.1) (Zymomonas mobilis) pir|A35260|A35260 alcohol dehydrogenase (EC 1.1.1.1) I - Zymomonas obilis	83	70	1053
95	2	708	2162	gi|506381	phospho-beta-glucosidase (Bacillus subtilis)	83	70	1455
137	1	68	694	gi|467391	initiation protein of replicaton (Bacillus subtilis)	83	77	627
140	4	3209	2742	gi|634107	kdpB (Escherichia coli)	83	65	468
142	3	3468	2989	gi|1212776	lumazine synthase (b-subunit) (Bacillus amyloliquefaciens)	83	69	480
161	12	5749	6696	gi|903307	ORF75 (Bacillus subtilis)	83	64	948
164	9	9880	11070	gi|49316	ORF2 gene product (Bacillus subtilis)	83	66	1191
164	14	14148	14546	gi|580902	ORF6 gene product (Bacillus subtilis)	83	60	399
170	2	3144	2467	gi|520844	orf4 (Bacillus subtilis)	83	64	678
186	2	2029	1370	gi|289284	cysteinyl-tRNA synthetase (Bacillus subtilis)	83	72	660
205	14	7822	7607	gi|216337	ORF for L30 ribosomal protein (Bacillus subtilis)	83	74	216
237	6	3683	4540	gi|1510488	imidazoleglycerol-phosphate synthase (cyclase) (Methanococcus jannaschii)	83	60	858
301	1	985	638	gi|467419	unknown (Bacillus subtilis)	83	65	348
302	4	1421	2743	gi|508979	GTP-binding protein (Bacillus subtilis)	83	68	1323
321	4	3933	3571	gi|39844	fumarase (citG) (aa 1-462) (Bacillus subtilis)	83	68	363
367	1	2	352	gi|1039479	ORFU (Lactococcus lactis)	83	54	351

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
387	1	3	662	gi|806281	DNA polymerase I [Bacillus stearothermophilus]	83	70	660
527	2	916	1566	gi|396259	protease [Staphylococcus epidermidis]	83	67	651
533	1	355	179	gi|142455	alanine dehydrogenase (EC 1.4.1.1) [Bacillus stearothermophilus] pir|B34261|B34261 alanine dehydrogenase (EC 1.4.1.1) - Bacillus tearothermophilus	83	66	177
536	4	1617	1438	gi|143366	adenylosuccinate lyase (PUR-B) [Bacillus subtilis] ,ir|C29326|WZBSDS adenylosuccinate lyase (EC 4.3.2.2) - Bacillus ubtilis	83	67	180
652	1	2	859	gi|520753	DNA topoisomerase I [Bacillus subtilis]	83	72	858
774	2	200	361	gi|1522665	M. jannaschii predicted coding region MJECL28 [Methanococcus jannaschii]	83	58	162
897	1	120	296	gi|1064807	ORTHININE AMINOTRANSFERASE [Bacillus subtilis]	83	76	177
1213	1	3	491	gi|289288	lexA [Bacillus subtilis]	83	67	489
2529	1	296	150	gi|143786	tryptophanyl-tRNA synthetase (EC 6.1.1.2) [Bacillus subtilis] pir|JT0481|YWBS tryptophan--tRNA ligase (EC 6.1.1.2) - Bacillus ubtilis	83	69	147
2973	1	649	326	gi|1109687	ProZ [Bacillus subtilis]	83	58	324
1009	1	728	366	gi|882532	ORF_o294 [Escherichia coli]	83	65	363
3035	2	45	305	gi|950062	hypothetical yeast protein 1 (Mycoplasma capricolum) pir|S48578|S48578 hypothetical protein - Mycoplasma capricolum SGC3) (fragment)	83	59	261
1906	1	67	309	gi|1353197	thioredoxin reductase [Eubacterium acidaminophilum]	83	61	243
4458	1	540	271	gi|397526	clumping factor [Staphylococcus aureus]	83	78	270
4570	1	444	223	gi|1022726	unknown [Staphylococcus haemolyticus]	83	74	222
4654	1	97	261	gi|1072419	glcB gene product [Staphylococcus carnosus]	83	79	165
16	2	295	1191	gi|153854	uvs402 protein [Streptococcus pneumoniae]	82	67	897
16	3	1193	1798	gi|153854	uvs402 protein [Streptococcus pneumoniae]	82	70	606
38	12	9644	8724	gi|1204400	N-acetylneuraminate lyase [Haemophilus influenzae]	82	58	921
42	4	988	2019	gi|841192	catalase [Bacteroides fragilis]	82	70	1032
51	6	2590	3489	gi|143607	sporulation protein [Bacillus subtilis]	82	69	900
56	11	12270	13925	gi|39431	oligo-1,6-glucosidase [Bacillus cereus]	82	60	1656
56	15	17673	18014	gi|467410	unknown [Bacillus subtilis]	82	66	342
61	2	881	3313	gi|143148	transfer RNA-Leu synthetase [Bacillus subtilis]	82	70	2433

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
82	7	9162	11318	gi|48240	elongation factor G (AA 1-691) (Thermus aquaticus thermophilus) ir|S15928|EFTWG translation elongation factor G - Thermus aquaticus p|P13551|EFG_THETH ELONGATION FACTOR G (EF-G).	82	64	2157
85	2	5470	3260	gi|143369	phosphoribosylformyl glycinamidine synthetase II (PUR-Q) (Bacillus ubtilis)	82	66	2211
102	6	3662	5380	gi|1256635	dihydroxy-acid dehydratase (Bacillus subtilis)	82	65	1719
117	4	3242	3493	pir|A47154|A471	orf1 5' of Ffh - Bacillus subtilis	82	53	252
128	6	4377	5933	gi|460258	phosphoglycerate mutase (Bacillus subtilis)	82	66	1557
129	2	1229	2182	gi|403373	glycerophosphoryl diester phosphodiesterase (Bacillus subtilis) pir|S37251|S37251 glycerophosphoryl diester phosphodiesterase - acillus subtilis	82	62	954
170	1	2	1441	gi|1377831	unknown (Bacillus subtilis)	82	67	1440
177	1	3	1094	gi|467386	thiophen and furan oxidation (Bacillus subtilis)	82	65	1092
184	4	3572	4039	gi|153566	ORF (19K protein) (Enterococcus faecalis)	82	59	468
189	8	4455	4225	gi|1001878	CspL protein (Listeria monocytogenes)	82	73	231
206	19	21366	20707	gi|473916	lipopeptide antibiotics iturin A (Bacillus subtilis) sp|P39144|LP14_BACSU LIPOPEPTIDE ANTIBIOTICS ITURIN A AND SURFACTIN IOSYNTHESIS PROTEIN.	82	50	660
221	2	805	1722	gi|517205	67 kDa Myosin-crossreactive streptococcal antigen (Streptococcus yogenes)	82	63	918
223	4	3866	3651	gi|439619	(Salmonella typhimurium IS200 insertion sequence from SARA17, artial.), gene product (Salmonella typhimurium)	82	69	216
260	3	5207	4296	gi|1161381	IcaB (Staphylococcus epidermidis)	82	61	912
315	3	4864	2855	gi|143397	quinol oxidase (Bacillus subtilis)	82	67	2010
321	10	8520	7945	gi|142981	ORF5; This ORF includes a region (aa23-103) containing a potential ron-sulphur centre homologous to a region of Rhodospirillum ruhrum nd Chromatium vinosum; putative (Bacillus stearothermophilus) pir|PQ0299|PQ0299 hypothetical protein 5 (gldA 3' region) -	82	62	576
331	3	1055	1342	gi|436574	ribosomal protein L1 (Bacillus subtilis)	82	71	288
370	2	262	618	gi|1303793	YqeL (Bacillus subtilis)	82	59	357
404	4	3053	4024	gi|1303821	YqfE (Bacillus subtilis)	82	68	972
405	4	4440	3073	gi|1303913	YqhX (Bacillus subtilis)	82	67	1368
436	3	4096	2864	gi|149521	tryptophan synthase beta subunit (Lactococcus lactis) pir|S35129|S35129 tryptophan synthase (EC 4.2.1.20) beta chain - actococcus lactis subsp. lactis	82	67	1233

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name ( )	% sim	% ident	length (nt)
441	4	3394	2573	gi|142952	glyceraldehyde-3-phosphate dehydrogenase (Bacillus tearothermophilus)	82	67	822
444	12	10415	11227	gi|1204354	spore germination and vegetative growth protein (Haemophilus influenzae)	82	67	813
446	1	3	191	gi|143387	aspartate transcarbamylase (Bacillus subtilis)	82	66	189
462	3	1007	1210	gi|142521	deoxyribodipyrimidine photolyase (Bacillus subtilis) pir|A37192|A37192 uvrB protein - Bacillus subtilis sp|P14951|UVRC_BACSU EXCINUCLEASE ABC SUBUNIT C.	82	64	204
537	1	1560	784	gi|853767	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (Bacillus ubtilis)	82	61	777
680	2	407	700	gi|426472	secE gene product (Staphylococcus carnosus)	82	69	294
724	2	565	386	gi|143373	phosphoribosyl aminoimidazole carboxy formyl ormyltransferase/inosine monophosphate cyclohydrolase (PUR-H(J)) Bacillus subtilis)	82	68	180
763	1	422	213	gi|467458	cell division protein (Bacillus subtilis)	82	35	210
818	1	564	283	gi|1064787	function unknown (Bacillus subtilis)	82	69	282
858	1	175	1176	gi|143043	uroporphyrinogen decarboxylase (Bacillus subtilis) pir|B47045|B47045 uroporphyrinogen decarboxylase (EC 4.1.1.37) - acillus subtilis	82	71	1002
895	1	3	599	gi|1027507	ATP binding protein (Borrelia burgdorferi)	82	72	597
939	1	10	399	gi|143795	transfer RNA-Tyr synthetase (Bacillus subtilis)	82	60	390
961	1	1	306	gi|577647	gamma-hemolysin (Staphylococcus aureus)	82	69	306
1192	1	307	155	gi|146974	NH3-dependent NAD synthetase (Escherichia coli)	82	71	153
1317	1	49	375	gi|407908	EIIscr (Staphylococcus xylosus)	82	72	327
1341	1	1	150	gi|39962	ribosomal protein L35 (AA 1-66) (Bacillus stearothermophilus) ir|S05347|R5BS35 ribosomal protein L35 - Bacillus earothermophilus	82	68	150
2990	2	567	349	gi|534855	ATPase subunit epsilon (Bacillus stearothermophilus) sp|P42009|ATPE_BACST ATP SYNTHASE EPSILON CHAIN (EC 3.6.1.34).	82	47	219
3024	1	45	224	gi|467402	unknown (Bacillus subtilis)	82	64	180
3045	1	276	139	gi|467335	ribosomal protein L9 (Bacillus subtilis)	82	60	138
3045	2	558	400	gi|467335	ribosomal protein L9 (Bacillus subtilis)	82	82	159
3091	1	474	238	gi|499335	secA protein (Staphylococcus carnosus)	82	78	237
3107	1	416	210	gi|546918	orfY 3' of comK (Bacillus subtilis, E26, Peptide Partial, 140 aa) pir|S43612|S43612 hypothetical protein Y - Bacillus subtilis sp|P40398|YHXD_BACSU HYPOTHETICAL PROTEIN IN COMK 3'REGION (ORFY) FRAGMENT).	82	64	207

TABLE 2

EP 0 786 519 A2

S. auraus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
4332	1	2	319	gi|42086	nitrate reductase alpha subunit (Escherichia coli) p|P09152|NARG_ECOLI RESPIRATORY NITRATE REDUCTASE 1 ALPHA CHAIN (EC 7.99.4). (SUB 2-1247)	82	75	318
23	3	3275	2574	gi|1199573	spsB (Sphingomonas sp.)	81	64	702
42	1	638	321	gi|466778	lysine specific permease (Escherichia coli)	81	59	318
48	5	4051	4350	gi|1045937	M. genitalium predicted coding region MG246 (Mycoplasma genitalium)	81	62	300
51	4	1578	2579	pir|S16649|S166	dciAC protein - Bacillus subtilis	81	55	1002
53	2	354	1494	gi|1303961	YqjJ (Bacillus subtilis)	81	67	1131
53	8	9419	7971	gi|146930	6-phosphogluconate dehydrogenase (Escherichia coli)	81	66	1449
54	9	10757	10119	gi|143016	permease (Bacillus subtilis)	81	65	639
54	10	13360	11786	gi|143015	gluconate kinase (Bacillus subtilis)	81	64	1575
57	17	13983	13366	pir|A25805|A258	L-lactate dehydrogenase (EC 1.1.1.27) - Bacillus subtilis	81	74	618
81	2	2708	2217	gi|1222302	NifU-related protein (Haemophilus influenzae)	81	54	492
86	1	745	374	gi|414017	ipa-93d gene product (Bacillus subtilis)	81	70	372
103	6	6438	4861	gi|971342	nitrate reductase beta subunit (Bacillus subtilis) sp|P42176|NARH_BACSU NITRATE REDUCTASE BETA CHAIN (EC 1.7.99.4).	81	64	1578
120	15	10845	12338	gi|1524392	GbsA (Bacillus subtilis)	81	67	1494
128	5	3676	4413	gi|143319	triose phosphate isomerase (Bacillus megaterium)	81	64	738
131	9	10308	9280	gi|299163	alanine dehydrogenase (Bacillus subtilis)	81	68	1029
143	6	6088	5471	gi|439619	(Salmonella typhimurium IS200 insertion sequence (from SARA17, artial.), gene product (Salmonella typhimurium)	81	61	618
169	1	43	825	gi|897795	30S ribosomal protein (Pediococcus acidilactici) sp|P49668|RS2_PEDAC 30S RIBOSOMAL PROTEIN S2.	81	65	783
230	1	450	226	gi|1125826	short region of weak similarity to tyrosine-protein kinase receptors in a fibronectin type III-like domain (Caenorhabditis elegans)	81	54	225
233	5	2000	2677	gi|467404	unknown (Bacillus subtilis)	81	63	678
241	2	3081	2149	gi|16510	succinate--CoA ligase (GDP-forming) (Arabidopsis thaliana) ir|S30579|S30579 succinate--CoA ligase (GDP-forming) (EC 6.2.1.4) pha chain - Arabidopsis thaliana (fragment)	81	69	933
256	1	1	981	pir|S09411|S094	spoIIIE protein - Bacillus subtilis	81	65	981
259	3	3752	2691	sp|P28367|RF2_B	PROBABLE PEPTIDE CHAIN RELEASE FACTOR 2 (RF-2) (FRAGMENT).	81	65	1062

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
275	2	1728	3581	gi|726480	L-glutamine-D-fructose-6-phosphate amidotransferase [Bacillus ubtilis]	81	68	1854
285	1	1466	735	gi|1204844	H. influenzae predicted coding region HI0594 [Haemophilus influenzae]	81	63	732
296	1	99	1406	gi|467328	adenylosuccinate synthetase [Bacillus subtilis]	81	67	1308
302	9	5590	5889	gi|147485	queA [Escherichia coli]	81	64	300
317	2	1137	1376	gi|154961	resolvase [Transposon Tn917]	81	5[illegible]	240
343	2	1034	1342	gi|405955	yeeD [Escherichia coli]	81	60	309
360	2	1404	2471	gi|1204570	aspartyl-tRNA synthetase [Haemophilus influenzae]	81	67	1068
364	5	6251	5706	gi|1204652	methylated-DNA--protein-cysteine methyltransferase [Haemophilus influenzae]	81	63	546
372	2	1707	1135	gi|467416	unknown [Bacillus subtilis]	81	65	573
392	1	43	603	pir|S09411|S094	spoIIIE protein - Bacillus subtilis	81	65	561
404	9	5252	6154	gi|606745	Bex [Bacillus subtilis]	81	65	903
426	2	1727	1119	gi|39453	Manganese superoxide dismutase [Bacillus caldotenax] ir|S22053|S22053 superoxide dismutase (EC 1.15.1.1) (Mn) - Bacillus ldotenax	81	66	609
480	7	[illegible]653	5889	pir|C37083|C370	hypothetical protein II (ompH 3' region) - Salmonella typhimurium (fragment)	81	57	237
625	3	1105	2070	gi|1262360	protein kinase PknB [Mycobacterium leprae]	81	56	966
754	2	504	1064	gi|1303902	YqhU [Bacillus subtilis]	81	71	561
842	1	86	430	gi|1405446	transketolase [Bacillus subtilis]	81	68	345
953	1	798	400	gi|1205429	dipeptide transport ATP-binding protein [Haemophilus influenzae]	81	57	399
961	2	252	401	gi|487686	synergohymenotropic toxin [Staphylococcus intermedius] pir|S44944|S44944 synergohymenotropic toxin - Staphylococcus ntermedius	81	72	150
1035	1	1	189	gi|1046138	M. genitalium predicted coding region MG423 [Mycoplasma genitalium]	81	43	189
1280	1	670	449	gi|559164	helicase [Autographa californica nuclear polyhedrosis virus] sp|P24307|V143_NPVAC HELICASE.	81	43	222
3371	1	68	241	gi|1322245	mevalonate pyrophosphate decarboxylase [Rattus norvegicus]	81	62	174
3715	1	475	239	gi|537137	ORF_f388 [Escherichia coli]	81	58	237
3908	1	2	325	gi|439619	[Salmonella typhimurium IS200 insertion sequence from SARA17, artial.], gene product [Salmonella typhimurium]	81	68	324
3940	1	3	401	gi|296464	ATPase [Lactococcus lactis]	81	69	399

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% identity	length (nt)
3954	1	1	318	gi|1224069	amidase (Moraxella catarrhalis)	81	68	318
4049	1	337	170	gi|603768	HutI protein, imidazolone-5-propionate hydrolase (Bacillus subtilis) gi|603768 HutI protein, imidazolone-5-propionate hydrolase Bacillus subtilis)	81	68	168
4209	1	1	324	gi|403373	glycerophosphoryl diester phosphodiesterase (Bacillus subtilis) pir|S37251|S37251 glycerophosphoryl diester phosphodiesterase - acillus subtilis	81	58	324
4371	1	627	322	gi|216677	indolepyruvate decarboxylase (Enterobacter cloacae) pir|S16013|S16013 indolepyruvate decarboxylase (EC 4.1.1.-) - nterobacter cloacae	81	72	306
4387	1	19	228	gi|460689	TVG (Thermoactinomyces vulgaris)	81	59	210
4191	1	581	306	gi|1524193	unknown (Mycobacterium tuberculosis)	81	67	276
4425	1	3	341	gi|143015	gluconate kinase (Bacillus subtilis)	81	66	339
9	1	1593	847	gi|1064786	function unknown (Bacillus subtilis)	80	62	747
17	1	544	311	gi|559164	helicase (Autographa californica nuclear polyhedrosis virus) sp|P24307|V143_NPVAC HELICASE.	80	40	234
45	2	1159	2448	gi|1109684	ProV (Bacillus subtilis)	80	63	1290
45	5	4032	4733	gi|1109687	ProZ (Bacillus subtilis)	80	55	702
54	8	10266	9502	gi|563952	gluconate permease (Bacillus licheniformis)	80	62	765
62	12	8852	7545	gi|854655	Na/H antiporter system (Bacillus alcalophilus)	80	62	1308
62	14	8087	8683	gi|559713	ORF (Homo sapiens)	80	68	597
67	16	13781	14122	gi|305002	ORF_f356 (Escherichia coli)	80	65	342
70	13	11495	10296	gi|1303995	YqkN (Bacillus subtilis)	80	64	1200
98	9	6336	7130	gi|467428	unknown (Bacillus subtilis)	80	68	795
98	10	7294	7833	gi|467430	unknown (Bacillus subtilis)	80	64	540
98	11	7820	8737	gi|467431	high level kasgamycin resistance (Bacillus subtilis)	80	61	918
109	16	14154	14813	gi|580875	ipa-57d gene product (Bacillus subtilis)	80	63	660
112	15	14294	16636	gi|1072361	pyruvate-formate-lyase (Clostridium pasteurianum)	80	65	2343
139	1	1448	726	gi|506699	CapC (Staphylococcus aureus)	80	58	723
139	2	2179	1448	gi|506698	CapB (Staphylococcus aureus)	80	59	732
174	4	3271	2870	gi|1146242	aspartate 1-decarboxylase (Bacillus subtilis)	80	61	402

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
177	3	2102	2842	gi|467385	unknown [Bacillus subtilis]	80	70	741
184	6	6124	5912	gi|161953	85-kDa surface antigen [Trypanosoma cruzi]	80	46	213
186	4	5368	3875	gi|289282	glutamyl-tRNA synthetase [Bacillus subtilis]	80	65	1494
205	30	15796	15140	gi|40103	ribosomal protein L4 [Bacillus stearothermophilus]	80	66	657
207	1	140	1315	gi|460259	enolase [Bacillus subtilis]	80	67	1176
211	3	1078	1590	gi|410131	ORFX7 [Bacillus subtilis]	80	61	513
235	2	1962	2255	gi|143797	valyl-tRNA synthetase [Bacillus stearothermophilus] sp|P11931|SYV_BACST VALYL-TRNA SYNTHETASE (EC 6.1.1.9) VALINE--TRNA LIGASE) (VALRS).	80	55	294
239	1	1	1263	gi|143000	proton glutamate symport protein [Bacillus stearothermophilus] pir|S26247|S26247 glutamate/aspartate transport protein - Bacillus tearothermophilus	80	59	1263
272	5	2724	2461	gi|709993	hypothetical protein [Bacillus subtilis]	80	54	264
301	3	1446	1111	gi|467418	unknown [Bacillus subtilis]	80	58	336
310	4	5697	4501	gi|1177686	acuC gene product [Staphylococcus xylosus]	80	67	1197
310	6	5258	7006	gi|348053	acetyl-CoA synthetase [Bacillus subtilis]	80	67	1749
310	7	7410	9113	gi|1103865	formyl-tetrahydrofolate synthetase [Streptococcus mutans]	80	67	1704
325	3	1114	1389	gi|310325	outer capsid protein [Rotavirus sp.]	80	40	276
337	1	1268	636	gi|537049	ORF_o470 [Escherichia coli]	80	55	633
374	2	929	1228	gi|1405448	YneF [Bacillus subtilis]	80	70	300
375	5	3062	3331	gi|467448	unknown [Bacillus subtilis]	80	68	270
388	1	267	587	gi|1064791	function unknown [Bacillus subtilis]	80	65	321
394	1	9	659	gi|304976	matches PS00017: ATP_GTP_A and PS00301: EFACTOR_GTP; similar to longation factor G, TetM/TetO tetracycline-resistance proteins Escherichia coli]	80	65	651
456	1	625	1263	gi|1146183	putative [Bacillus subtilis]	80	65	639
475	1	1	654	gi|288269	beta-fructofuranosidase [Staphylococcus xylosus]	80	66	654
544	2	1449	2240	gi|529754	speC [Streptococcus pyogenes]	80	50	792
622	4	1623	1871	gi|1483545	unknown [Mycobacterium tuberculosis]	80	65	249
719	1	1	1257	gi|1064791	function unknown [Bacillus subtilis]	80	68	1257
739	1	107	838	gi|666983	putative ATP binding subunit [Bacillus subtilis]	80	61	732

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name ( )	% sim	% idont	length (nt)
745	2	581	414	gi|1511600	coenzyme PQQ synthesis protein III (Methanococcus jannaschii)	80	61	168
822	1	17	679	gi|410141	ORFX17 (Bacillus subtilis)	80	68	663
827	2	991	836	gi|1205301	leukotoxin secretion ATP-binding protein (Haemophilus influenzae)	80	54	156
1044	1	3	149	gi|60632	vp2 (Marburg virus)	80	55	147
1220	2	571	413	pir|A61072|EPSG	gallidermin precursor - Staphylococcus gallinarum	80	7%	159
2519	1	75	275	gi|147556	dpj (Escherichia coli)	80	45	201
2947	1	503	279	gi|1184680	polynucleotide phosphorylase (Bacillus subtilis)	80	62	225
3120	1	2	226	gi|517205	67 kDa Myosin-crossreactive streptococcal antigen (Streptococcus yogenes)	80	65	225
3191	1	294	148	gi|151259	HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas mevalonii) pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	80	59	147
3560	2	285	434	gi|217130	photosystem I core protein B (Synechococcus vulcanus)	80	70	150
3655	1	47	346	gi|415855	deoxyribose aldolase (Mycoplasma hominis)	80	56	300
3658	2	324	584	gi|551531	2-nitropropane dioxygenase (Williopsis saturnus)	80	54	261
3769	1	798	400	gi|1339950	large subunit of NADH-dependent glutamate synthase (Plectonema boryanum)	80	68	399
3781	1	692	348	gi|166412	NADH-glutamate synthase (Medicago sativa)	80	62	345
3988	1	48	287	gi|1204696	fructose-permease IIBC component (Haemophilus influenzae)	80	69	240
4030	1	571	287	gi|1009366	Respiratory nitrate reductase (Bacillus subtilis)	80	60	285
4092	1	547	275	gi|1370207	orf6 (Lactobacillus sake)	80	69	273
4103	1	680	342	gi|39956	IIGlc (Bacillus subtilis)	80	65	339
4231	1	692	348	gi|289287	UDP-glucose pyrophosphorylase (Bacillus subtilis)	80	65	345
4265	1	595	299	gi|603768	HutI protein, imidazolone-5-propionate hydrolase (Bacillus subtilis) gi|603768 HutI protein, imidazolone-5-propionate hydrolase Bacillus subtilis)	80	63	297
4504	1	498	250	gi|1339950	large subunit of NADH-dependent glutamate synthase (Plectonema boryanum)	80	68	249
2	6	5998	6798	g |535351	CodY (Bacillus subtilis)	79	63	801
4	7	8295	7051	gi|603768	HutI protein, imidazolone-5-propionate hydrolase (Bacillus subtilis) gi|603768 HutI protein, imidazolone-5-propionate hydrolase Bacillus subtilis)	79	64	1245
25	6	5273	5515	pir|A36728|A367	acyl carrier protein - Rhizobium meliloti	79	65	243

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
59	2	1173	1424	gi|147923	threonine dehydratase 2 (EC 4.2.1.16) [Escherichia coli]	79	75	252
60	1	1	204	gi|666115	orf1 upstream of glucose kinase (Staphylococcus xylosus) pir|S52351|S52351 hypothetical protein 1 - Staphylococcus xylosus	79	60	204
81	1	3002	1590	gi|466882	pps1; B1496_C2_189 [Mycobacterium leprae]	79	64	1413
85	7	7023	6505	gi|143364	phosphoribosyl aminoimidazole carboxylase I (PUR-E) [Bacillus ubtilis]	79	60	519
89	6	5660	4554	gi|144906	product homologous to E.coli thioredoxin reductase: J.Biol.Chem. 1988) 263:9015-9019, and to F52a protein of alkyl hydroperoxide eductase from S.typhimurium: J.Biol.Chem. (1990) 265:10535-10540; pen reading frame A [Clostridium pasteurianum]	79	35	1107
102	11	7489	8571	gi|143093	ketol-acid reductoisomerase [Bacillus subtilis] sp|P37253|ILVC_BACSU KETOL-ACID REDUCTOISOMERASE (EC 1.1.1.86) ACETOHYDROXY-ACID ISOMEROREDUCTASE) (ALPHA-KETO-BETA-HYDROXYLACIL EDUCTOISOMERASE).	79	64	1083
102	14	11190	12563	gi|149428	putative [Lactococcus lactis]	79	65	1374
127	9	7792	9372	gi|458688	PrfC/RF3 [Dichelobacter nodosus]	79	68	1581
139	3	2540	1983	gi|506697	CapA (Staphylococcus aureus)	79	55	558
144	2	1644	1156	gi|1498296	peptide methionine sulfoxido reductase [Streptococcus pneumoniae]	79	47	489
148	2	529	1098	gi|467457	hypoxanthine-guanine phosphoribosyltransferase [Bacillus subtilis] gi|467457 hypoxanthine-guanine phosphoribosyltransferase [Bacillus ubtilis]	79	59	570
150	1	965	591	gi|755602	unknown [Bacillus subtilis]	79	61	375
176	1	1039	587	gi|297874	fructose-bisphosphate aldolase [Staphylococcus carnosus] pir|A49943|A49943 fructose-bisphosphate aldolase (EC 4.1.2.13) - taphylococcus carnosus (strain TM300)	79	65	453
186	7	7584	6874	gi|1314298	ORF5; putative Sms protein; similar to Sms proteins from Haemophilus influenzae and Escherichia coli [Listeria monocytogenes]	79	64	711
205	16	8887	8498	gi|1044980	ribosomal protein L18 [Bacillus subtilis]	79	70	390
211	1	1	519	gi|1303994	YqkH [Bacillus subtilis]	79	62	519
223	2	4183	2801	gi|488430	alcohol dehydrogenase 2 [Entamoeba histolytica]	79	60	1383
243	8	8915	7896	gi|580883	ipa-88d gene product [Bacillus subtilis]	79	60	1020
279	4	3721	4329	gi|413930	ipa-6d gene product [Bacillus subtilis]	79	59	609
300	1	11	1393	gi|403372	glycerol 3-phosphate permease [Bacillus subtilis]	79	62	1383
307	3	2930	1935	gi|950062	hypothetical yeast protein 1 [Mycoplasma capricolum] pir|S48578|S48578 hypothetical protein - Mycoplasma capricolum SGC3) (fragment)	79	60	996

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
352	6	10106	8886	gi|216854	P47K (Pseudomonas chlororaphis)	79	59	1221
412	1	1153	578	gi|143177	putative (Bacillus subtilis)	79	51	576
481	3	621	1124	gi|786163	Ribosomal Protein L10 (Bacillus subtilis)	79	66	504
516	1	702	352	gi|805090	NisF (Lactococcus lactis)	79	48	351
525	2	2457	1426	gi|143371	phosphoribosyl aminoimidazole synthetase (PUR-M) (Bacillus subtilis) pir|H29326|AJBSCL phosphoribosylformylglycinamidine cyclo-ligase EC 6.3.3.1) - Bacillus subtilis	79	61	1032
538	4	3448	2825	gi|1370207	orf6 (Lactobacillus sake)	79	67	624
570	1	2	421	gi|476160	arginine permease substrate-binding subunit (Listeria monocytogenes)	79	61	420
645	8	2663	3241	gi|153898	transport protein (Salmonella typhimurium)	79	62	579
683	1	75	374	gi|1064795	function unknown (Bacillus subtilis)	79	62	300
816	3	4700	3987	gi|1407784	orf-1; novel antigen (Staphylococcus aureus)	79	62	714
2929	1	3	401	gi|1524397	glycine betaine transporter OpuD (Bacillus subtilis)	79	61	399
2937	1	357	202	pir|S52915|S529	nitrate reductase alpha chain - Bacillus subtilis (fragment)	79	58	156
2940	1	768	385	gi|149429	putative (Lactococcus lactis)	79	72	384
2946	1	570	286	gi|143267	2-oxoglutarate dehydrogenase (odhA; EC 1.2.4.2) (Bacillus subtilis)	79	61	285
2999	1	3	212	gi|710020	nitrite reductase (nirB) (Bacillus subtilis)	79	59	210
3022	1	514	332	gi|450686	3-phosphoglycerate kinase (Thermotoga maritima)	79	61	183
3064	1	3	314	gi|1204436	pyruvate formate-lyase (Haemophilus influenzae)	79	60	312
3083	1	2	220	gi|1149662	hypD gene product (Clostridium perfringens)	79	56	219
3126	1	701	411	gi|1339950	large subunit of NADH-dependent glutamate synthase (Plectonema boryanum)	79	55	291
3181	1	607	326	gi|1339950	large subunit of NADH-dependent glutamate synthase (Plectonema boryanum)	79	59	282
3345	1	3	476	gi|871784	Clp-like ATP-dependent protease binding subunit (Bos taurus)	79	63	474
3718	1	536	270	pir|C36889|C368	leuB protein, inactive - Lactococcus lactis subsp. lactis (strain IL1403)	79	71	267
3724	2	159	401	gi|1009366	Respiratory nitrate reductase (Bacillus subtilis)	79	64	243
3836	1	608	312	gi|1524193	unknown (Mycobacterium tuberculosis)	79	65	297
3941	1	2	334	gi|415855	deoxyribose aldolase (Mycoplasma hominis)	79	54	333
4113	1	3	341	gi|143015	gluconate kinase (Bacillus subtilis)	79	63	339

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name ( )	% sim	% ident	length (nt)
4501	1	406	209	gi|1022726	unknown [Staphylococcus haemolyticus]	79	66	198
4612	1	2	238	gi|460689	TVG [Thermoactinomyces vulgaris]	79	58	237
2	1	2	1213	gi|520753	DNA topoisomerase I (Bacillus subtilis]	78	64	1212
8	2	2266	1220	gi|216151	DNA polymerase (gene L; ttg start codon) [Bacteriophage SPO2] gi|579197 SPO2 DNA polymerase (aa 1-648) [Bacteriophage SPO2] pir|A21498|DJBPS2 DNA-directed DNA polymerase (EC 2.7.7.7) - phage PO2	78	72	1047
9	2	1340	1089	gi|1064787	function unknown [Bacillus subtilis]	78	57	252
32	8	6803	7702	gi|146974	NH3-dependent NAD synthetase [Escherichia coli]	78	63	900
36	4	2941	3138	gi|290503	glutamate permease [Escherichia coli]	78	53	198
53	15	17684	16221	gi|1303941	YqiV [Bacillus subtilis]	78	58	1464
57	14	10520	12067	gi|1072418	glcA gene product [Staphylococcus carnosus]	78	65	1548
66	7	6798	5812	gi|1212729	YqhJ [Bacillus subtilis]	78	67	987
67	4	4029	4376	gi|466612	nikA [Escherichia coli]	78	71	348
91	9	10058	10942	gi|467380	stage 0 sporultion [Bacillus subtilis]	78	50	885
102	12	8574	10130	gi|149426	putative [Lactococcus lactis]	78	61	1557
112	6	3540	4463	gi|854234	cymG gene product [Klebsiella oxytoca]	78	56	924
124	2	1888	1061	gi|405622	unknown [Bacillus subtilis]	78	60	828
130	3	1805	2260	gi|1256636	putative [Bacillus subtilis]	78	71	456
133	1	751	377	gi|168060	lamB [Emericella nidulans]	78	59	375
166	4	7125	6163	gi|451216	Mannosephosphate Isomerase [Streptococcus mutans]	78	63	963
186	1	1586	795	gi|289284	cysteinyl-tRNA synthetase [Bacillus subtilis]	78	63	792
195	4	2749	2315	gi|1353874	unknown [Rhodobacter capsulatus]	78	58	435
199	3	4279	3623	gi|143525	succinate dehydrogenase cytochrome b-558 subunit [Bacillus subtilis] pir|A29843|DEBSSC succinate dehydrogenase (EC 1.3.99.1) cytochrome 558 - Bacillus subtilis	78	57	657
199	4	7209	5557	gi|142521	deoxyribodipyrimidine photolyase [Bacillus subtilis] pir|A37192|A37192 uvrB protein - Bacillus subtilis sp|P14951|UVRC_BACSU EXCINUCLEASE ABC SUBUNIT C.	78	62	1653
223	3	3831	3523	gi|139596	[Escherichia coli IS200 insertion sequence from ECOR63, partial.], ene product [Escherichia coli]	78	47	309

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name ( )	% sim	% ident	length (nt)
299	4	1865	2149	gi|467439	temperature sensitive cell division (Bacillus subtilis)	78	62	285
321	9	7734	7315	gi|142979	ORF3 is homologous to an ORF downstream of the spoT gene of E.coli; RF3 (Bacillus stearothermophilus)	78	55	420
352	4	3714	3944	gi|349050	actin 1 (Pneumocystis carinii)	78	42	231
352	5	7592	6093	gi|903587	NADH dehydrogenase subunit 5 (Bacillus subtilis) sp|P39755|NDHF_BACSU NADH DEHYDROGENASE SUBUNIT 5 (EC 1.6.5.3) NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5).	78	58	1500
376	1	2	583	gi|551693	dethiobiotin synthase (Bacillus sphaericus)	78	34	582
424	2	1595	1768	gi|1524117	alpha-acetolactate decarboxylase (Lactococcus lactis)	78	68	174
450	1	1914	988	gi|1030068	NAD(P)H oxidoreductase, isoflavone reductase homologue (Solanum tuberosum)	78	63	927
558	1	762	562	gi|1511588	bifunctional protein (Methanococcus jannaschii)	78	60	201
670	3	1152	1589	gi|1122759	unknown (Bacillus subtilis)	78	64	438
714	1	64	732	gi|143460	37 kd minor sigma factor (rpoF, sigB; ttg start codon) (Bacillus ubtilis)	78	57	669
814	1	3	368	gi|1377833	unknown (Bacillus subtilis)	78	59	366
981	1	1381	692	gi|143802	GerC2 (Bacillus subtilis)	78	64	690
995	2	978	727	gi|296947	uridine kinase (Escherichia coli)	78	64	252
1045	1	3	401	gi|1407784	orf-1; novel antigen (Staphylococcus aureus)	78	61	399
1163	2	368	186	gi|410117	diaminopimelate decarboxylase (Bacillus subtilis)	78	54	183
2191	1	794	399	gi|215098	excisionase (Bacteriophage 154a)	78	65	396
2933	1	2	181	gi|1204436	pyruvate formate-lyase (Haemophilus influenzae)	78	73	180
3041	2	129	317	gi|624632	GltL (Escherichia coli)	78	53	189
3581	1	105	401	gi|763186	3-ketoacyl-coA thiolase (Saccharomyces cerevisiae)	78	55	297
3709	1	3	230	gi|460689	TVG (Thermoactinomyces vulgaris)	78	58	228
3974	1	528	265	gi|558839	unknown (Bacillus subtilis)	78	65	264
3980	1	3	401	gi|39956	IIGlc (Bacillus subtilis)	78	62	399
4056	1	647	354	gi|1256635	dihydroxy-acid dehydratase (Bacillus subtilis)	78	55	294
4114	1	630	316	pir|S09372|S093	hypothetical protein - Trypanosoma brucei	78	62	315
4185	1	3	179	gi|1339950	large subunit of NADH-dependent glutamate synthase (Plectonema boryanum)	78	58	177

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
4235	1	655	329	gi|558839	unknown [Bacillus subtilis]	78	60	327
4352	1	541	302	gi|603768	HutI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis] gi|603768 HutI protein, imidazolone-5-propionate hydrolase Bacillus subtilis]	78	63	240
4368	1	612	307	gi|1353678	heavy-metal transporting P-type ATPase [Proteus mirabilis]	78	59	306
4461	1	428	216	gi|1276841	glutamate synthase (GOGAT) [Porphyra purpurea]	78	36	213
4530	1	474	238	gi|39956	IIGlc [Bacillus subtilis]	78	65	237
3	2	2969	2073	gi|1109684	ProV [Bacillus subtilis]	77	56	897
12	2	2426	1965	gi|467335	ribosomal protein L9 [Bacillus subtilis]	77	59	462
27	1	2	388	gi|1212728	YqhI [Bacillus subtilis]	77	63	387
39	2	590	1252	gi|40054	phenylalanyl-tRNA synthetase beta subunit (AA 1-804) [Bacillus btilis]	77	60	663
42	6	2704	2931	gi|606241	30S ribosomal subunit protein S14 [Escherichia coli] sp|P02370|RS14_ECOLI 30S RIBOSOMAL PROTEIN S14. (SUB 2-101)	77	65	228
46	18	15459	16622	gi|297798	mitochondrial formate dehydrogenase precursor [Solanum tuberosum] pir|JQ2272|JQ2272 formate dehydrogenase (EC 1.2.1.2) precursor, itochondrial - potato	77	55	1164
100	4	4562	4002	gi|1340128	ORF1 [Staphylococcus aureus]	77	54	561
102	8	5378	5713	gi|1311482	acetolactate synthase [Thermus aquaticus]	77	57	336
109	7	4742	5383	gi|710637	Unknown [Bacillus subtilis]	77	56	642
117	1	2	1228	gi|1237015	ORF4 [Bacillus subtilis]	77	53	1227
124	10	8323	7688	gi|405819	thymidine kinase [Bacillus subtilis]	77	63	636
147	3	1146	985	gi|849027	hypothetical 15.9-kDa protein [Bacillus subtilis]	77	37	162
152	10	7354	7953	gi|1205583	spermidine/putrescine transport ATP-binding protein [Haemophilus influenzae]	77	55	600
169	2	1004	1282	gi|473825	'elongation factor EF-Ts' [Escherichia coli]	77	58	279
184	2	380	1147	gi|216314	esterase [Bacillus stearothermophilus]	77	60	768
189	7	3296	3868	gi|853809	ORF3 [Clostridium perfringens]	77	48	573
193	1	132	290	gi|1303788	YqeH [Bacillus subtilis]	77	54	159
195	8	8740	8414	gi|1499620	M. jannaschii predicted coding region MJ0798 [Methanococcus jannaschii]	77	44	327
205	8	5428	5204	gi|216340	ORF for adenylate kinase [Bacillus subtilis]	77	61	225

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
205	29	14795	14502	gi|786155	Ribosomal Protein L23 (Bacillus subtilis)	77	62	294
211	5	1908	2084	gi|410132	ORFX8 (Bacillus subtilis)	77	47	177
217	5	3478	4416	gi|496254	fibronectin/fibrinogen-binding protein (Streptococcus pyogenes)	77	54	939
232	1	267	998	gi|1407784	orf-1; novel antigen (Staphylococcus aureus)	77	57	732
233	2	1819	1346	gi|467408	unknown (Bacillus subtilis)	77	61	474
243	3	2661	2299	gi|516155	unconventional myosin (Sus scrofa)	77	32	363
299	1	68	769	gi|467436	unknown (Bacillus subtilis)	77	54	702
301	4	1468	1283	gi|950071	ATP-bind. pyrimidine kinase (Mycoplasma capricolum) pir|S48605|S48605 hypothetical protein - Mycoplasma capricolum SGC3) (fragment)	77	48	186
302	5	2741	3211	gi|508980	pheB (Bacillus subtilis)	77	57	471
302	7	3835	4863	gi|147783	ruvB protein (Escherichia coli)	77	60	1029
307	9	5402	4797	gi|1070015	protein-dependent (Bacillus subtilis)	77	60	606
312	1	99	1391	gi|143165	malic enzyme (EC 1.1.1.38) (Bacillus stearothermophilus) pir|A33307|DEBSXS malate dehydrogenase oxaloacetate-decarboxylating) (EC 1.1.1.38) - Bacillus tearothermophilus	77	62	1293
312	2	1541	2443	gi|1399855	carboxyltransferase beta subunit (Synechococcus PCC7942)	77	58	903
321	5	5666	4596	gi|39844	fumarase (citG) (aa 1-462) (Bacillus subtilis)	77	65	1071
354	1	47	568	gi|1154634	YmaB (Bacillus subtilis)	77	57	522
365	1	2	1021	gi|143374	phosphoribosyl glycinamide synthetase (PUR-D; gtg start codon) Bacillus subtilis)	77	62	1020
374	1	1	708	gi|1405446	transketolase (Bacillus subtilis)	77	61	708
385	1	1128	565	gi|533099	endonuclease III (Bacillus subtilis)	77	63	564
392	2	594	1940	gi|556014	UDP-N-acetyl muramate-alanine ligase (Bacillus subtilis) sp|P40778|MURC_BACSU UDP-N-ACETYLMURAMATE--ALANINE LIGASE (EC .3.2.8) (UDP-N- ACETYLMURANOYL-L-ALANINE SYNTHETASE) (FRAGMENT).	77	65	1347
405	5	4079	3570	gi|1303912	YqhW (Bacillus subtilis)	77	64	510
487	4	1302	1472	gi|432427	ORF1 gene product (Acinetobacter calcoaceticus)	77	48	171
522	1	2	562	pir|A01179|SYBS	tyrosine--tRNA ligase (EC 6.1.1.1) - Bacillus stearothermophilus	77	63	561

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match Acession	match gene name	% sim	% ident	length (nt)
523	2	1587	1351	gi|1387979	44% identity over 302 residues with hypothetical protein from Synechocystis sp, accession D64006_CD; expression induced by environmental stress; some similarity to glycosyl transferases; two potential membrane-spanning helices [Bacillus subtil	77	48	237
536	2	983	612	gi|143366 -	adenylosuccinate lyase (PUR-B) [Bacillus subtilis] pir|C29326|WZBSDS adenylosuccinate lyase (EC 4.3.2.2) - Bacillus ubtilis	77	61	372
548	2	339	872	gi|143387	aspartate transcarbamylase [Bacillus subtilis]	77	56	534
597	1	2	481	gi|904198	hypothetical protein [Bacillus subtilis]	77	33	480
633	2	1747	1313	gi|387577	ORF1A [Bacillus subtilis]	77	64	435
642	1	85	360	gi|46971	epiP gene product [Staphylococcus epidermidis]	77	61	276
659	1	125	1219	gi|1072381	glutamyl-aminopeptidase [Lactococcus lactis]	77	62	1095
670	4	1587	1820	gi|1122760	unknown [Bacillus subtilis]	77	58	234
789	1	2	391	gi|1377823	aminopeptidase [Bacillus subtilis]	77	65	390
815	1	10	573	gi|1303861	YqgN [Bacillus subtilis]	77	49	564
899	1	1	225	gi|1204H44	H. influenzae predicted coding region HI0594 [Haemophilus influenzae]	77	55	225
1083	1	3	188	gi|460828	B969 [Saccharomyces cerevisiae]	77	66	186
1942	1	415	209	gi|160047	p101/acidic basic repeat antigen [Plasmodium falciparum] pir|A29232|A29232 101K malaria antigen precursor - Plasmodium alciparum (strain Camp)	77	38	207
2559	1	1	171	gi|1499034	M. jannaschii predicted coding region MJ0255 [Methanococcus jannaschii]	77	61	171
2933	2	243	401	gi|42370	pyruvate formate-lyase (AA 1-760) [Escherichia coli] ir|S01788|S01788 formate C-acetyltransferase (EC 2.3.1.54) - cherichia coli	77	72	159
2966	1	56	292	gi|1524397	glycine betaine transporter OpuD [Bacillus subtilis]	77	45	237
2976	1	614	309	gi|40003	oxoglutarate dehydrogenase (NADP+) [Bacillus subtilis] p|P23129|ODO1_BACSU 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (EC 2.4.2) (ALPHA- KETOGLUTARATE DEHYDROGENASE).	77	60	306
2979	2	678	400	gi|1204354	spore germination and vegetative growth protein [Haemophilus influenzae]	77	61	279
2988	1	601	377	gi|438465	Probable operon with orfF. Possible alternative initiation codon, ases 2151-2153. Homology with acetyltransferases.; putative Bacillus subtilis]	77	55	225
2990	1	331	167	gi|142562	ATP synthase epsilon subunit [Bacillus megaterium] pir|B28599|PWBSEM H+- transporting ATP synthase (EC 3.6.1.34) psilon chain - Bacillus megaterium	77	63	165
3032	1	3	389	gi|488430	alcohol dehydrogenase 2 [Entamoeba histolytica]	77	56	387
3057	1	1	195	gi|468764	mocR gene product [Rhizobium meliloti]	77	50	195

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
4008	1	726	400	gi|603768	HutI protein, imidazolone-5-propionate hydrolase (Bacillus subtilis) gi|603768 HutI protein, imidazolone-5-propionate hydrolase Bacillus subtilis	77	52	327
4048	1	703	386	gi|216278	gramicidin S synthetase 1 (Bacillus brevis)	77	55	318
4110	1	3	368	pir|S52915|S529	nitrate reductase alpha chain - Bacillus subtilis (fragment)	77	61	366
4115	1	1	348	gi|517205	67 kDa Myosin-crossreactive streptococcal antigen (Streptococcus yogenes)	77	65	348
4225	1	590	297	gi|1322245	mevalonate pyrophosphate decarboxylase (Rattus norvegicus)	77	60	294
4611	2	494	327	gi|508979	GTP-binding protein (Bacillus subtilis)	77	57	168
4668	1	361	182	pir|S52915|S529	nitrate reductase alpha chain - Bacillus subtilis (fragment)	77	61	180
25	1	2	1627	gi|1150620	HmsA (Streptococcus pneumoniae)	76	58	1626
38	5	1488	2537	pir|A43577|A435	regulatory protein pfoR - Clostridium perfringens	76	57	1050
52	5	2962	4041	gi|1161061	dioxygenase (Methylobacterium extorquens)	76	62	1080
56	20	27389	27955	gi|467402	unknown (Bacillus subtilis)	76	56	567
57	15	12046	12219	gi|1206040	weak similarity to keratin (Caenorhabditis elegans)	76	40	174
91	2	1062	2261	gi|475715	acetyl coenzyme A acetyltransferase (thiolase) (Clostridium cetobutylicum)	76	57	1200
98	2	818	1624	gi|467422	unknown (Bacillus subtilis)	76	62	807
98	5	2965	3228	gi|897793	y98 gene product (Pediococcus acidilactici)	76	52	264
98	8	5922	6326	gi|467427	methionyl-tRNA synthetase (Bacillus subtilis)	76	53	405
104	3	1322	1885	gi|216151	DNA polymerase (gene L; ttg start codon) (Bacteriophage SPO2) gi|579197 SPO2 DNA polymerase (aa 1-648) (Bacteriophage SPO2) pir|A21498|DJBPS2 DNA-directed DNA polymerase (EC 2.7.7.7) - phage PO2	76	63	564
124	9	8134	7055	gi|853776	peptide chain release factor 1 (Bacillus subtilis) pir|S55437|S55437 peptide chain release factor 1 - Bacillus ubtilis	76	58	1080
164	5	2832	3311	gi|1204976	prolyl-tRNA synthetase (Haemophilus influenzae)	76	53	480
168	2	2617	1841	gi|1177253	putative ATP-binding protein of ABC-type (Bacillus subtilis)	76	58	777
189	2	163	888	gi|467384	unknown (Bacillus subtilis)	76	63	726
235	3	2253	3518	gi|142936	folyl-polyglutamate synthetase (Bacillus subtilis) pir|B40646|B40646 folC - Bacillus subtilis	76	53	1266
236	1	335	925	gi|1146197	putative (Bacillus subtilis)	76	54	591
237	8	5323	5541	gi|1279261	F13G3.6 (Caenorhabditis elegans)	76	47	219

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
263	5	5490	4585	gi|1510348	dihydrodipicolinate synthase [Methanococcus jannaschii]	76	49	906
304	3	1051	1794	gi|666982	putative membrane spanning subunit [Bacillus subtilis] pir|S52382|S52382 probable membrane spanning protein - Bacillus ubtilis	76	60	744
312	4	3611	4624	gi|143312	6-phospho-1-fructokinase (gtg start codon; EC 2.7.1.11) [Bacillus tearothermophilus]	76	56	1014
343	1	2	1036	gi|405956	yeeE [Escherichia coli]	76	59	1035
347	1	409	1701	gi|396304	acetylornithine deacetylase [Escherichia coli]	76	72	1293
358	1	672	1907	gi|1146215	39.0% identity to the Escherichia coli S1 ribosomal protein; putative [Bacillus subtilis]	76	58	1236
371	1	1	222	gi|537084	alternate gene name mgt; CG Site No. 497 [Escherichia coli] pir|S56468|S56468 mgtA protein - Escherichia coli	76	61	222
379	4	4331	4858	gi|143268	dihydrolipoamide transsuccinylase (odhB; EC 2.3.1.61) [Bacillus ubtilis]	76	61	528
404	5	4022	4492	gi|1303823	YqfG [Bacillus subtilis]	76	60	471
411	1	2	307	gi|486025	ORF YKL027w [Saccharomyces cerevisiae]	76	55	306
472	3	4356	2854	gi|1405464	AlsT [Bacillus subtilis]	76	57	1503
546	1	273	995	gi|153821	streptococcal pyrogenic exotoxin type C (speC) precursor Streptococcus pyogenes]	76	36	723
588	1	1054	557	gi|1002520	MutS [Bacillus subtilis]	76	61	498
591	1	16	735	gi|885934	ClpB [Synechococcus sp.]	76	44	720
602	2	175	798	gi|1486422	OppD homologue [Rhizobium sp.]	76	52	624
619	2	547	290	gi|330613	major capsid protein [Human cytomegalovirus]	76	47	258
660	4	2568	3302	gi|904199	hypothetical protein [Bacillus subtilis]	76	55	735
677	1	452	228	gi|40177	spoOF gene product [Bacillus subtilis]	76	58	225
962	1	24	206	gi|142443	adenylosuccinate synthetase [Bacillus subtilis] sp|P29726|PURA_BACSU ADENYLOSUCCINATE SYNTHETASE (EC 6.3.4.4) IMP--ASPARTATE LIGASE).	76	67	183
978	1	1158	580	gi|1511333	M. jannaschii predicted coding region MJ1322 [Methanococcus jannaschii]	76	56	579
997	1	486	244	gi|467154	No definition line found [Mycobacterium leprae]	76	38	243
1563	1	529	266	gi|1303984	YqkG [Bacillus subtilis]	76	52	264
2184	1	361	182	gi|506706	CapJ [Staphylococcus aureus]	76	38	180
2572	1	1	387	gi|153898	transport protein [Salmonella typhimurium]	76	65	387

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
2942	1	29	400	gi|710020	nitrite reductase (nirB) [Bacillus subtilis]	76	59	372
2957	1	377	216	gi|1511251	hypothetical protein (SP:P42404) [Methanococcus jannaschii]	76	47	162
2980	1	554	279	gi|1405464	AlsT [Bacillus subtilis]	76	53	276
3015	1	649	326	gi|408115	ornithine acetyltransferase [Bacillus subtilis]	76	61	324
3124	1	13	174	gi|882705	ORF_o401 [Escherichia coli]	76	65	162
3179	1	3	161	gi|168477	ferredoxin-dependant glutamate synthase [Zea mays] pir|A38596|A38596 glutamate synthase (ferredoxin) (EC 1.4.7.1) - aize	76	53	159
3789	1	2	379	gi|39956	IIGlc [Bacillus subtilis]	76	55	378
3892	1	3	314	gi|1510398	ferripyochelin binding protein [Methanococcus jannaschii]	76	52	312
3928	1	798	400	gi|143016	permease [Bacillus subtilis]	76	59	399
4159	1	757	386	sp|P80544|MRSP_	METHICILLIN-RESISTANT SURFACE PROTEIN (FRAGMENTS).	76	66	372
4204	1	17	331	gi|296464	ATPase [Lactococcus lactis]	76	56	315
4398	1	494	249	gi|987255	Menkes disease gene [Homo sapiens]	76	48	246
4506	1	2	313	gi|216746	D-lactate dehydrogenase [Lactobacillus plantarum]	76	47	312
4546	1	477	247	gi|1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	76	61	231
4596	1	379	191	gi|560027	cellulose synthase [Acetobacter xylinum]	76	70	189
4	5	5257	4337	gi|882532	ORF_o294 [Escherichia coli]	75	59	921
6	1	164	952	gi|40960	OTCase [Escherichia coli]	75	56	789
12	3	5935	3944	gi|467336	unknown [Bacillus subtilis]	75	57	1992
23	18	18272	17310	gi|1296433	O-acetylserine sulfhydrylase B [Alcaligenes eutrophus]	75	55	963
25	3	2356	3393	gi|1502419	PlsX [Bacillus subtilis]	75	56	1038
36	8	5765	6037	gi|1256517	unknown [Schizosaccharomyces pombe]	75	45	273
46	13	11186	12058	gi|48972	nitrate transporter [Synechococcus sp.]	75	46	873
51	7	3474	3677	gi|143607	sporulation protein [Bacillus subtilis]	75	61	204
53	16	16850	16590	gi|143402	recombination protein (ttg start codon). [Bacillus subtilis] gi|1303923 RecN [Bacillus subtilis]	75	51	261
74	3	3572	2568	gi|1204847	ornithine carbamoyltransferase [Haemophilus influenzae]	75	61	1005

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
85	3	4628	3930	gi|143368	phosphoribosylformyl glycinamidine synthetase I (PUR-L; gtg start odon) [Bacillus subtilis]	75	63	699
85	5	5588	4878	gi|143367	phosphoribosyl aminoidazole succinocarboxamide synthetase (PUR-C; tg start codon) [Bacillus subtilis]	75	55	711
85	8	6625	7530	gi|1303916	YqiA [Bacillus subtilis]	75	53	906
87	3	2340	3590	gi|1064813	homologous to sp:PHOR_BACSU [Bacillus subtilis]	75	56	1251
87	6	6084	6896	gi|1064810	function unknown [Bacillus subtilis]	75	61	813
108	2	1844	1503	gi|1001824	hypothetical protein [Synechocystis sp.]	75	51	342
110	3	1748	3727	gi|1147593	putative ppGpp synthetase [Streptomyces coelicolor]	75	55	1980
110	7	4151	5252	gi|1177251	clwD gene product [Bacillus subtilis]	75	75	900
120	14	11266	10649	gi|1524394	ORF-2 upstream of gbsAB operon [Bacillus subtilis]	75	55	618
121	5	2050	4221	gi|1154632	NrdE [Bacillus subtilis]	75	54	2172
124	1	283	143	gi|405622	unknown [Bacillus subtilis]	75	58	141
128	1	81	1139	gi|143316	[gap] gene products [Bacillus megaterium]	75	48	1059
130	8	5760	5903	gi|1256654	54.8% identity with Neisseria gonorrhoeae regulatory protein PilB; putative [Bacillus subtilis]	75	62	144
136	2	4480	3185	gi|467403	seryl-tRNA synthetase [Bacillus subtilis]	75	54	1296
161	10	5439	5798	gi|1001195	hypothetical protein [Synechocystis sp.]	75	55	360
172	4	3819	2995	gi|755153	ATP-binding protein [Bacillus subtilis]	75	52	825
179	1	2024	1107	gi|143037	porphobilinogen deaminase [Bacillus subtilis]	75	58	918
195	10	9529	9374	sp|P25745|YCFD_	HYPOTHETICAL PROTEIN IN PURB 5'REGION (ORF-15) (FRAGMENT).	75	60	156
200	4	2605	4596	gi|142440	ATP-dependent nuclease [Bacillus subtilis]	75	56	1992
206	3	6900	5620	gi|1256135	YbbP [Bacillus subtilis]	75	53	1281
216	2	159	389	gi|1052800	unknown [Schizosaccharomyces pombe]	75	58	231
229	1	29	847	gi|1205958	branched chain aa transport system II carrier protein [Haemophilus influenzae]	75	49	819
230	2	518	1714	gi|971337	nitrite extrusion protein [Bacillus subtilis]	75	53	1197
231	1	2240	1122	gi|1002521	MutL [Bacillus subtilis]	75	54	1119
233	3	1314	1859	gi|467405	unknown [Bacillus subtilis]	75	59	546

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
269	1	325	164	gi|1511246	methyl coenzyme M reductase system, component A2 (Methanococcus jannaschii)	75	50	162
292	1	1389	772	gi|1511604	M. jannaschii predicted coding region MJ1651 (Methanococcus jannaschii)	75	46	618
304	4	1773	2261	gi|1205328	surfactin (Haemophilus influenzae)	75	55	489
312	3	2437	3387	gi|285621	undefined open reading frame (Bacillus stearothermophilus)	75	62	951
312	5	4622	6403	gi|1041097	Pyruvate Kinase (Bacillus psychrophilus)	75	57	1782
319	1	353	877	gi|1212728	YqhI (Bacillus subtilis)	75	54	525
320	5	4321	5031	gi|1070361	OMP decarboxylase (Lactococcus lactis)	75	56	711
320	6	5010	5642	gi|143394	OMP-PRPP transferase (Bacillus subtilis)	75	60	633
337	4	1519	2088	gi|487433	citrate synthase II (Bacillus subtilis)	75	58	570
394	2	669	1271	gi|304976	matches PS00017: ATP_GTP_A and PS00301: EFACTOR_GTP; similar to longation factor G, TetM/TetO tetracycline-resistance proteins Escherichia coli)	75	51	603
423	1	127	570	gi|1183839	unknown (Pseudomonas aeruginosa)	75	59	444
433	2	1603	1929	gi|149211	acetolactate synthase (Klebsiella pneumoniae)	75	63	327
446	2	176	1540	gi|312441	dihydroorotase (Bacillus caldolyticus)	75	62	1365
486	1	494	249	gi|1149682	potF gene product (Clostridium perfringens)	75	55	246
496	1	3	794	gi|143582	spoIIIEA protein (Bacillus subtilis)	75	59	792
498	2	824	1504	gi|143328	phoP protein (put.); putative (Bacillus subtilis)	75	47	681
499	2	1061	1624	gi|1387979	44% identity over 302 residues with hypothetical protein from Synechocystis sp, accession D64006_CD; expression induced by environmental stress; some similarity to glycosyl transferases; two potential membrane-spanning helices (Bacillus subtil	75	51	564
568	1	641	453	pir|JC4110|JC41	triacylglycerol lipase (EC 3.1.1.3) 2 - Mycoplasma mycoides subsp. mycoides (SGC3)	75	50	189
613	2	430	233	gi|330993	tegument protein (Saimiriine herpesvirus 2)	75	75	198
623	1	1	525	gi|529754	speC (Streptococcus pyogenes)	75	43	525
642	5	1809	2474	gi|1176401	EpiG (Staphylococcus epidermidis)	75	51	666
646	2	454	657	gi|172442	ribonuclease P (Saccharomyces cerevisiae)	75	37	204
657	1	3	347	gi|882561	ORF_o236 (Escherichia coli)	75	47	345
750	1	1662	832	gi|46971	epiP gene product (Staphylococcus epidermidis)	75	57	831

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name ( )	% sim	% idunt	length (nt)
754	1	2	481	gi|1303901	YqhT [Bacillus subtilis]	75	57	480
763	2	563	393	gi|1205145	multidrug resistance protein [Haemophilus influenzae]	75	51	171
775	1	961	482	pir|B36889|B368	leuA protein, inactive - Lactococcus lactis subsp. lactis (strain IL1403)	75	63	480
793	1	1	180	gi|143316	(gap) gene products [Bacillus megaterium]	75	57	180
800	1	318	160	gi|509411	NFRA protein [Azorhizobium caulinodans]	75	34	159
811	1	1117	560	gi|143434	Rho Factor [Bacillus subtilis]	75	60	558
940	1	493	329	gi|1276985	arginase [Bacillus caldovelox]	75	50	165
971	2	37	252	gi|1001373	hypothetical protein [Synechocystis sp.]	75	58	216
1059	1	384	232	gi|726480	L-glutamine-D-fructose-6-phosphate amidotransferase [Bacillus ubtilis]	75	67	153
1109	2	219	374	gi|143331	alkaline phosphatase regulatory protein [Bacillus subtilis] pir|A27650|A27650 regulatory protein phoR - Bacillus subtilis sp|P23545|PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN HOR (EC 2.7.3.-).	75	53	156
1268	1	271	137	gi|304135	ornithine acetyltransferase [Bacillus stearothermophilus] sp|Q07908|ARGJ_BACST GLUTAMATE N-ACETYLTRANSFERASE (EC 2.3.1.35) ORNITHINE ACETYLTRANSFERASE) (ORNITHINE TRANSACETYLASE) (OATASE) / MINO-ACID ACETYLTRANSFERASE (EC 2.3.1.1) (N-ACETYLGLUTAMATE YNTHA	75	63	135
1500	1	324	163	gi|1205488	excinuclease ABC subunit B [Haemophilus influenzae]	75	57	162
1529	1	798	400	gi|1002521	MutL [Bacillus subtilis]	75	54	399
3010	1	770	387	gi|1204435	pyruvate formate-lyase activating enzyme [Haemophilus influenzae]	75	54	384
3105	1	1	180	gi|1041097	Pyruvate Kinase [Bacillus psychrophilus]	75	57	180
3117	1	45	212	gi|899317	peptide synthetase module [Microcystis aeruginosa] pir|S49111|S49111 probable amino acid activating domain - icrocystis aeruginosa (fragment) (SUB 144-528)	75	42	168
3139	2	139	345	gi|145294	adenine phosphoribosyl-transferase [Escherichia coli]	75	66	207
3880	1	618	310	gi|1009366	Respiratory nitrate reductase [Bacillus subtilis]	75	58	309
3911	1	48	401	gi|433991	ATP synthase subunit beta [Bacillus subtilis]	75	68	354
3957	1	2	379	pir|B36889|B368	3-isopropylmalate dehydratase (EC 4.2.1.33) chain leuC - Lactococcus lactis subsp. lactis (strain IL1403)	75	65	378
4005	1	5	259	gi|216746	D-lactate dehydrogenase [Lactobacillus plantarum]	75	48	255
4080	1	73	333	gi|415855	deoxyribose aldolase [Mycoplasma hominis]	75	59	261

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4111	1	1	339	gi|149435	putative [Lactococcus lactis]	75	57	339
4136	1	602	303	gi|450688	hsdM gene of EcoprrI gene product [Escherichia coli] pir|S38437|S38437 hsdM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	75	56	300
4144	1	668	336	gi|48972	nitrate transporter [Synechococcus sp.]	75	49	333
4237	1	664	374	gi|1339950	large subunit of NADH-dependent glutamate synthase [Plectoneme boryanum]	75	55	291
4306	2	73	318	gi|294260	major surface glycoprotein [Pneumocystis carinii]	75	68	246
4343	1	715	359	gi|1204652	methylated-DNA--protein-cysteine methyltransferase [Haemophilus influenzae]	75	52	357
4552	1	620	312	gi|296464	ATPase [Lactococcus lactis]	75	55	309
38	9	5776	6126	gi|443793	NupC [Escherichia coli]	74	50	351
50	8	6910	6221	gi|1239988	hypothetical protein [Bacillus subtilis]	74	55	690
56	9	10770	12221	gi|1000451	TreP [Bacillus subtilis]	74	57	1452
64	2	2266	1622	gi|41015	aspartate-tRNA ligase [Escherichia coli]	74	57	645
66	6	5063	4848	gi|1212729	YqhJ [Bacillus subtilis]	74	47	216
67	18	14334	14897	gi|1510631	endoglucanase [Methanococcus jannaschii]	74	52	564
102	15	12561	13136	gi|149429	putative [Lactococcus lactis]	74	67	576
102	16	13121	14419	gi|149435	putative [Lactococcus lactis]	74	57	1299
108	4	4873	3902	gi|39478	ATP binding protein of transport ATPases [Bacillus firmus] ir|S15486|S15486 ATP-binding protein - Bacillus firmus p|P26946|YATR_BACFI HYPOTHETICAL ATP-BINDING TRANSPORT PROTEIN.	74	59	972
116	5	8574	7093	gi|1205430	dipeptide transport system permease protein [Haemophilus influenzae]	74	49	1482
120	7	4342	4803	gi|146970	ribonucleoside triphosphate reductase [Escherichia coli] pir|A47331|A47331 anaerobic ribonucleotide reductase - Escherichia oli	74	58	462
121	7	5961	6581	gi|1107528	ttg start [Campylobacter coli]	74	51	621
128	3	2320	3531	gi|143318	phosphoglycerate kinase [Bacillus megaterium]	74	57	1212
130	7	5237	5791	gi|1256653	DNA-binding protein [Bacillus subtilis]	74	60	555
136	3	6745	5150	gi|143076	histidase [Bacillus subtilis]	74	58	1596
145	2	664	1368	gi|407773	devA gene product [Anabaena sp.]	74	45	705
152	1	552	277	gi|1377833	unknown [Bacillus subtilis]	74	54	276

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
164	10	11064	11375	gi|580900	ORF3 gene product [Bacillus subtilis]	74	52	312
175	2	3109	2624	gi|642656	unknown [Rhizobium meliloti]	74	34	486
175	9	6064	5612	gi|854656	Na/H antiporter system ORF2 [Bacillus alcalophilus]	74	46	453
195	11	11346	10339	gi|1204430	hypothetical protein (SP:P25745) [Haemophilus influenzae]	74	55	1008
205	17	9619	9059	gi|1044979	ribosomal protein L6 [Bacillus subtilis]	74	64	561
236	7	5574	6710	gi|1146207	putative [Bacillus subtilis]	74	63	1137
241	3	4521	3334	gi|694121	malate thiokinase [Methylobacterium extorquens]	74	52	1188
246	6	3305	2799	gi|467374	single strand DNA binding protein [Bacillus subtilis]	74	64	507
249	4	6551	5313	gi|1524397	glycine betaine transporter OpuD [Bacillus subtilis]	74	55	1239
261	7	4389	4081	gi|809542	CbrB protein [Erwinia chrysanthemi]	74	42	309
278	6	5714	4665	gi|1204872	ATP-binding protein [Haemophilus influenzae]	74	54	1050
309	1	1220	666	gi|1205579	hypothetical protein (GB:U14003_302) [Haemophilus influenzae]	74	53	555
315	2	1473	862	gi|143398	quinol oxidase [Bacillus subtilis]	74	57	612
320	1	1	1065	gi|143389	glutaminase of carbamyl phosphate synthetase [Bacillus subtilis] pir|E39845|E39845 carbamoyl-phosphate synthase glutamine-hydrolyzing) (EC 6.3.5.5), pyrimidine-repressible, small hain - Bacillus subtilis	74	60	1065
380	2	382	1128	gi|534857	ATPase subunit a [Bacillus stearothermophilus]	74	56	747
405	2	1742	1311	gi|1303915	YqhZ [Bacillus subtilis]	74	65	432
433	5	2503	3270	gi|473902	alpha-acetolactate synthase [Lactococcus lactis]	74	56	768
452	1	1	942	gi|413982	ipa-58r gene product [Bacillus subtilis]	74	52	942
461	1	3	1193	gi|558494	homoserine dehydrogenase [Bacillus subtilis]	74	51	1191
461	2	1174	1407	gi|40211	threonine synthase (thrC) (AA 1-352) [Bacillus subtilis] ir|A25364|A25364 threonine synthase (EC 4.2.99.2) - Bacillus btilis	74	56	234
462	2	402	734	gi|142520	thioredoxin [Bacillus subtilis]	74	62	333
478	1	574	320	gi|1499005	glycyl-tRNA synthetase [Methanococcus jannaschii]	74	52	255
501	2	739	1740	gi|217040	acid glycoprotein [Streptococcus pyogenes]	74	58	1002
551	2	4083	2791	gi|143040	glutamate-1-semialdehyde 2,1-aminotransferase [Bacillus subtilis] pir|D42728|D42728 glutamate-1-semialdehyde 2,1-aminomutase (EC .4.3.8) - Bacillus subtilis	74	51	1293

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
573	1	1	477	gi|1006605	hypothetical protein [Synechocystis sp.]	74	45	477
596	2	1780	1298	gi|1301853	YqqF [Bacillus subtilis]	74	55	483
618	2	2924	1758	gi|1146237	21.4% of identity to trans-acting transcription factor of Sacharomyces cerevisiae; 25% of identity to sucrose synthase of Zea mays; putative [Bacillus subtilis]	74	55	1167
659	2	1269	1595	gi|1072380	ORF3 [Lactococcus lactis]	74	62	327
724	1	373	188	gi|143374	phosphoribosyl glycinamide synthetase (PUR-D; gtg start codon) Bacillus subtilis]	74	58	186
743	2	604	1209	gi|153833	ORF1; putative [Streptococcus parasanguis]	74	50	606
816	1	2	259	gi|143458	ORF V [Bacillus subtilis]	74	47	258
989	2	443	724	gi|1303994	YqkM [Bacillus subtilis]	74	46	282
1106	1	1	492	gi|46970	epiD gene product [Staphylococcus epidermidis]	74	54	492
1135	2	373	528	gi|413948	ipa-24d gene product [Bacillus subtilis]	74	48	156
1234	1	817	452	gi|495245	recJ gene product [Erwinia chrysanthemi]	74	36	366
2586	1	2	238	gi|1149701	sbcC gene product [Clostridium perfringens]	74	62	237
2959	1	798	400	gi|1405454	aconitase [Bacillus subtilis]	74	60	399
2962	1	650	363	gi|450686	3-phosphoglycerate kinase [Thermotoga maritima]	74	58	288
2983	1	3	191	gi|1303893	YqhL [Bacillus subtilis]	74	56	189
3018	1	2	223	gi|143040	glutamate-1-semialdehyde 2,1-aminotransferase [Bacillus subtilis] pir|D42728|D42728 glutamate-1-semialdehyde 2,1-aminomutase (EC .4.3.8) - Bacillus subtilis	74	56	222
3038	1	510	256	pir|S52915|S529	nitrate reductase alpha chain - Bacillus subtilis (fragment)	74	57	255
3062	1	374	189	gi|1107528	ttg start [Campylobacter coli]	74	51	186
4035	1	184	360	gi|1022725	unknown [Staphylococcus haemolyticus]	74	64	177
4045	1	607	305	gi|1510977	M. jannaschii predicted coding region MJ0938 [Methanococcus jannaschii]	74	41	303
4283	1	471	304	gi|520844	orf4 [Bacillus subtilis]	74	58	168
4449	1	3	221	gi|580910	peptide-synthetase ORF1 [Bacillus subtilis]	74	54	219
4587	1	458	231	gi|1370207	orf6 [Lactobacillus sake]	74	59	228

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
4603	1	29	214	gi|146208	glutamate synthase large subunit (EC 2.6.1.53) [Escherichia coli] pir|A29617|A29617 glutamate synthase (NADPH) (EC 1.4.1.13) large hain - Escherichia coli	74	60	186
4670	1	366	184	gi|1256135	YbbF [Bacillus subtilis]	74	61	183
5	10	7953	7162	gi|143727	putative [Bacillus subtilis]	73	42	792
11	2	2454	1372	gi|166338	dihydroorotate dehydrogenase [Agrocybe aegerita]	73	55	1083
14	1	2024	1020	gi|143373	phosphoribosyl aminoimidazole carboxy formyl ormyltransferase/inosine monophosphate cyclohydrolase (PUR-H(J)) Bacillus subtilis]	73	54	1005
23	5	5426	4635	gi|1468939	meso-2,3-butanediol dehydrogenase (D-acetoin forming) (Klebsiella pneumoniae]	73	58	792
23	17	17379	16360	gi|297060	ornithine cyclodeaminase [Rhizobium meliloti]	73	37	1020
29	2	692	1273	gi|467442	stage V sporulation [Bacillus subtilis]	73	54	582
31	5	6467	4914	gi|414000	ipa-76d gene product [Bacillus subtilis]	73	55	1554
37	8	8658	7402	gi|429259	pepT gene product [Bacillus subtilis]	73	59	1257
37	9	7738	7562	gi|168367	alpha-isopropylmalate isomerase (puc.); putative [Rhizomucor ircinelloides]	73	52	177
38	7	3931	4896	gi|405885	yeiN [Escherichia coli]	73	58	966
44	6	5041	4238	gi|580895	unknown [Bacillus subtilis]	73	53	804
44	11	7767	8306	gi|42009	moaB gene product [Escherichia coli]	73	50	540
45	3	2439	3080	gi|1109685	ProW [Bacillus subtilis]	73	47	642
54	13	14036	13794	gi|413931	ipa-7d gene product [Bacillus subtilis]	73	61	243
59	4	1430	2248	gi|147923	threonine dehydratase 2 (EC 4.2.1.16) [Escherichia coli]	73	53	819
65	1	1458	730	gi|677944	AppF [Bacillus subtilis]	73	56	729
80	2	1375	860	gi|580932	murD gene product [Bacillus subtilis]	73	53	516
102	13	10124	11179	gi|580891	3-isopropylmalate dehydrogenase (AA 1 - 365) [Bacillus subtilis] pir|A26522|A26522 3-isopropylmalate dehydrogenase (EC 1.1.1.85) - acillus subtilis	73	55	1056
109	2	3493	2600	gi|1510849	M. jannaschii predicted coding region MJ0775 [Methanococcus jannaschii]	73	40	894
120	8	4782	5756	gi|146970	ribonucleoside triphosphate reductase [Escherichia coli] pir|A47331|A47331 anaerobic ribonucleotide reductase - Escherichia oli	73	56	975
120	9	5726	6223	gi|1204333	anaerobic ribonucleoside-triphosphate reductase [Haemophilus influenzae]	73	62	498

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
132	5	4151	4363	gi|871048	HPSR2 - heavy chain potential motor protein [Giardia intestinalis]	73	43	213
140	6	5952	4324	gi|634107	kdpB [Escherichia coli]	73	59	1629
142	6	7060	5919	gi|410125	riboG gene product [Bacillus subtilis]	73	57	1122
149	4	1866	1717	gi|460892	heparin binding protein-44, HBP-44 (mice, Peptide, 360 aa) pir|JX0281|JX0281 heparin-binding protein-44 precursor - mouse gi|220434 ORF [Mus musculus] (SUB 2-360)	73	53	150
158	1	1	1431	gi|882504	ORF_f560 [Escherichia coli]	73	57	1431
174	6	5352	4525	gi|1146240	ketopantoate hydroxymethyltransferase [Bacillus subtilis]	73	55	828
175	8	5537	5178	gi|854657	Na/H antiporter system ORF3 [Bacillus alcalophilus]	73	56	360
186	5	6593	5493	gi|467477	unknown [Bacillus subtilis]	73	48	1101
249	6	6283	5729	gi|1524397	glycine betaine transporter OpuD [Bacillus subtilis]	73	56	555
265	4	1873	2280	gi|39848	U3 [Bacillus subtilis]	73	41	408
270	1	328	582	gi|780461	220 kDa polyprotein [African swine fever virus]	73	53	255
278	4	4283	3618	gi|1208965	hypothetical 23.3 kd protein [Escherichia coli]	73	49	666
279	3	4984	3593	gi|1185288	isochorismate synthase [Bacillus subtilis]	73	58	1392
291	4	1207	1575	gi|1511440	glutamine--fructose-6-phosphate transaminase [Methanococcus jannaschii]	73	63	369
299	2	735	1166	gi|467437	unknown [Bacillus subtilis]	73	58	432
299	5	2050	3234	gi|467439	temperature sensitive cell division [Bacillus subtilis]	73	53	1185
334	1	1237	728	gi|536655	ORF YBR244w [Saccharomyces cerevisiae]	73	43	510
336	2	1827	1036	gi|790943	urea amidolyase [Bacillus subtilis]	73	51	792
374	3	1389	1874	gi|1405451	YneJ [Bacillus subtilis]	73	55	486
433	4	1916	2554	gi|473902	alpha-acetolactate synthase [Lactococcus lactis]	73	54	639
509	2	1795	1028	gi|467483	unknown [Bacillus subtilis]	73	56	768
513	1	1709	918	gi|1146220	NAD+ dependent glycerol-3-phosphate dehydrogenase [Bacillus subtilis]	73	56	792
533	2	239	733	gi|1510605	hypothetical protein (SP:P42297) [Methanococcus jannaschii]	73	44	495
546	2	1148	2815	gi|41748	hsdM protein (AA 1-520) [Escherichia coli]	73	52	1668
549	1	762	382	gi|1314847	CinA [Bacillus subtilis]	73	57	381
567	1	1346	675	gi|410137	ORFX13 [Bacillus subtilis]	73	58	672

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
716	2	654	1112	gi|1256623	exodeoxyribonuclease (Bacillus subtilis)	73	56	459
772	1	3	677	gi|142010	Shows 70.2% similarity and 48.6% identity to the EnvM protein of almonella typhimurium (Anabaena sp.)	73	57	675
774	1	3	209	gi|409286	bmrU (Bacillus subtilis)	73	52	207
782	1	1	402	gi|143320	(gap) gene products (Bacillus megaterium)	73	56	402
789	2	451	762	gi|1063246	low homology to P14 protein of Heamophilus influenzar and 14.2 kDa protein of Escherichia coli (Bacillus subtilis)	73	56	312
796	1	3	911	gi|453754	ABC transporter (Bacillus subtilis)	73	58	909
806	3	1209	949	gi|143786	tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Bacillus subtilis) pir|JT0481|YWBS tryptophan--tRNA ligase (EC 6.1.1.2) - Bacillus ubtilis	73	51	261
816	2	4839	3097	gi|41748	hsdM protein (AA 1-520) (Escherichia coli)	73	52	1743
839	1	798	400	gi|886906	argininosuccinate synthetase (Streptomyces clavuligerus) pir|S57659|S57659 argininosuccinate synthase (EC 6.3.4.5) - treptomyces clavuligerus	73	59	399
857	1	3	290	gi|348052	acetoin utilization protein (Bacillus subtilis)	73	50	288
1008	1	790	398	gi|40100	rodC (tagJ) polypeptide (AA 1-746) (Bacillus subtilis) ir|S06049|S06049 rodC protein - Bacillus subtilis p|P13485|TAGF_BACSU TECHOIC ACID BIOSYNTHESIS PROTEIN F.	73	41	393
1018	1	1	213	gi|529357	No definition line found (Caenorhabditis elegans) sp|P46975|STT3_CAEEL OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT ONOLOG.	73	53	213
1031	1	3	491	gi|142706	comG1 gene product (Bacillus subtilis)	73	51	489
1174	1	395	204	gi|1149513	alpha3a subunit of laminin 5 (Homo sapiens)	73	60	192
1175	1	655	329	gi|473817	'ORF' (Escherichia coli)	73	57	327
1187	1	3	209	gi|580870	ipa-37d qoxA gene product (Bacillus subtilis)	73	52	207
1206	1	72	245	gi|144816	formyltetrahydrofolate synthetase (FTHFS) (ttg start codon) (EC .3.4.3) (Moorella thermoacetica)	73	43	174
1454	1	423	241	gi|1213253	unknown (Schizosaccharomyces pombe)	73	53	183
1469	1	517	260	gi|1303787	YqeG (Bacillus subtilis)	73	55	258
1761	1	374	189	gi|9135	Mst26Aa gene product (Drosophila simulans)	73	34	186
1849	1	467	243	gi|162307	DNA topoisomerase II (Trypanosoma cruzi)	73	60	225
2055	1	2	400	gi|559381	P47K protein (Rhodococcus erythropolis)	73	34	399
2556	1	2	244	gi|145925	fecB (Escherichia coli)	73	62	243

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
2947	2	549	400	gi|1184680	polynucleotide phosphorylase [Bacillus subtilis]	73	51	150
2956	1	746	375	gi|143397	quinol oxidase [Bacillus subtilis]	73	58	372
3037	1	655	329	gi|143091	acetolactate synthase [Bacillus subtilis]	73	55	327
3115	1	385	194	gi|323866	overlapping out-of-phase protein [Eggplant mosaic virus] sp|P20129|V70K_EPMV 70 KD PROTEIN.	73	53	192
3603	2	700	527	gi|1419521	glutaryl-CoA dehydrogenase precursor [Mus musculus]	73	48	174
3743	1	798	400	gi|450688	hsdM gene of EcoprrI gene product [Escherichia coli] pir|S38437|S38437 hsdM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli [SUB 40-520]	73	54	399
3752	1	640	359	gi|1524193	unknown [Mycobacterium tuberculosis]	73	59	282
3852	1	2	181	gi|216746	D-lactate dehydrogenase [Lactobacillus plantarum]	73	68	180
3914	1	475	239	pir|S13490|S134	Hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - Chicken (fragment)	73	53	237
3914	2	570	343	gi|528991	unknown [Bacillus subtilis]	73	38	228
4069	1	2	316	gi|40003	oxoglutarate dehydrogenase (NADP+) [Bacillus subtilis] p|P23129|ODO1_BACSU 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (EC 2.4.2) (ALPHA- KETOGLUTARATE DEHYDROGENASE).	73	55	315
4165	1	715	365	gi|1439521	glutaryl-CoA dehydrogenase precursor [Mus musculus]	73	48	351
4196	1	1	177	gi|809660	deoxyribose-phosphate aldolase [Bacillus subtilis] pir|S49455|S49455 deoxyribose-phosphate aldolase (EC 4.1.2.4) - acillus subtilis	73	60	177
4202	1	572	378	gi|528991	unknown [Bacillus subtilis]	73	38	195
4314	1	2	193	gi|436797	N-acyl-L-amino acid amidohydrolase [Bacillus stearothermophilus] sp|P37112|AHA_BACST N-ACYL-L-AMINO ACID AMIDOHYDROLASE (EC .5.1.14) (AMINOACYLASE).	73	47	192
4393	1	3	263	gi|216267	ORF2 [Bacillus megaterium]	73	47	261
35	2	903	1973	gi|1146196	phosphoglycerate dehydrogenase [Bacillus subtilis]	72	53	1071
38	22	19094	17877	gi|602031	similar to trimethylamine DH [Mycoplasma capricolum] pir|S49950|S49950 probable trimethylamine dehydrogenase (EC .5.99.7) - Mycoplasma capricolum (SGC3) (fragment)	72	54	1218
38	23	18134	19162	gi|413968	ipa-44d gene product [Bacillus subtilis]	72	54	1029
44	19	11895	12953	gi|516272	unknown [Bacillus subtilis]	72	49	1059
48	7	6248	7117	gi|43499	pyruvate synthase [Halobacterium halobium]	72	49	870
50	7	6563	5691	gi|1205399	proton glutamate symport protein [Haemophilus influenzae]	72	53	873

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
53	9	10521	9259	gi|1303956	YqjE (Bacillus subtilis)	72	52	1263
56	23	29549	29995	gi|467471	unknown (Bacillus subtilis)	72	47	447
69	4	5298	4123	gi|1354775	pfoS/R (Treponema pallidum)	72	46	1176
69	5	4377	4982	gi|904198	hypothetical protein (Bacillus subtilis)	72	43	606
73	1	2	856	gi|142997	glycerol uptake facilitator (Bacillus subtilis)	72	59	855
98	13	9371	10258	gi|467435	unknown (Bacillus subtilis)	72	50	888
127	1	1	1593	gi|217144	alanine carrier protein (thermophilic bacterium PS3) pir|A45111|A45111 alanine transport protein - thermophilic acterium PS-3	72	56	1593
131	1	5197	2600	gi|153952	polymerase III polymerase subunit (dnaE) (Salmonella typhimurium) pir|A45915|A45915 DNA-directed DNA polymerase (EC 2.7.7.7) III lpha chain - Salmonella typhimurium	72	53	2598
141	4	1040	1978	gi|1405446	transketolase (Bacillus subtilis)	72	54	939
149	8	2819	2535	gi|606234	secY (Escherichia coli)	72	44	285
149	17	5472	5245	gi|1304472	DNA polymerase (Unidentified phycodnavirus clone OTU4)	72	55	228
154	1	1	210	gi|1205620	ferritin like protein (Haemophilus influenzae)	72	40	210
155	1	2207	1320	gi|391610	farnesyl diphosphate synthase (Bacillus stearothermophilus) pir|JX0257|JX0257 geranyltranstransferase (EC 2.5.1.10) - Bacillus tearothermophilus	72	57	888
180	1	2	328	gi|433630	A180 (Saccharomyces cerevisiae)	72	62	327
184	3	1145	3553	gi|1205110	virulence associated protein homolog (Haemophilus influenzae)	72	49	2409
195	2	1923	1279	gi|1001730	hypothetical protein (Synechocystis sp.)	72	45	645
206	13	14646	15869	gi|1064807	ORTHININE AMINOTRANSFERASE (Bacillus subtilis)	72	50	1224
209	2	462	932	gi|1206666	hypothetical protein (GB:X73124_53) (Haemophilus influenzae)	72	60	471
215	2	764	522	gi|881513	insulin receptor homolog (Drosophila melanogaster) pir|S57245|S57245 insulin receptor homolog - fruit fly (Drosophila elanogaster) (SUB 46-2146)	72	63	243
224	1	2	790	gi|949974	sucrose repressor (Staphylococcus xylosus)	72	54	789
233	1	1526	765	gi|1408493	homologous to SwissProt:YIDA_ECOLI hypothetical protein (Bacillus subtilis)	72	52	762
240	1	220	1485	gi|537049	ORF_o470 (Escherichia coli)	72	52	1266
245	1	3	1340	gi|1206578	hypothetical protein (GB:U06949_1) (Haemophilus influenzae)	72	46	1338

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
259	2	2108	1245	gi|1340128	ORF1 [Staphylococcus aureus]	72	59	864
304	2	285	1094	gi|1205330	glutamine-binding periplasmic protein [Haemophilus influenzae]	72	52	810
307	10	5326	5039	gi|1070015	protein-dependent [Bacillus subtilis]	72	53	288
315	1	517	260	gi|143399	quinol oxidase [Bacillus subtilis]	72	55	258
316	11	9622	9308	gi|1206445	hypothetical protein (SP:P27857) [Haemophilus influenzae]	72	58	315
337	3	926	1609	gi|487433	citrate synthase II [Bacillus subtilis]	72	55	684
364	7	12538	10493	gi|1510643	ferrous iron transport protein B [Methanococcus jannaschii]	72	53	2046
409	2	340	1263	gi|1402944	orfRM1 gene product [Bacillus subtilis]	72	49	924
441	3	2177	1590	gi|312379	highly conserved among eubacteria [Clostridium acetobutylicum] pir|S34312|S34312 hypothetical protein V - Clostridium cetobutylicum	72	48	588
453	6	2654	2505	pir|S00601|BXSA	antibacterial protein 3 - Staphylococcus haemolyticus	72	70	150
460	1	2	625	gi|1016162	ABC transporter subunit [Cyanophora paradoxa]	72	51	624
463	1	3253	1628	gi|666014	The polymorphysm (RFLP) of this gene is associated with usceptibility to essential hypertension. The SA gene product has light homology to acetyl-CoA synthetase [Homo sapiens]	72	60	1626
480	4	3047	3466	gi|433992	ATP synthase subunit epsilon [Bacillus subtilis]	72	53	420
502	1	1086	586	gi|310859	ORF2 [Synechococcus sp.]	72	50	501
519	1	81	1184	gi|1303704	YrkE [Bacillus subtilis]	72	54	1104
559	1	3	746	gi|1107530	ceuD gene product [Campylobacter coli]	72	56	744
575	1	1142	573	gi|1303866	YqgS [Bacillus subtilis]	72	56	570
671	1	2	592	gi|1204497	protein-export membrane protein [Haemophilus influenzae]	72	44	591
679	2	295	1251	gi|563258	virulence-associated protein E [Dichelobacter nodosus]	72	52	957
687	2	295	957	gi|1146214	44% identical amino acids with the Escherichia coli smba supress; putative [Bacillus subtilis]	72	49	663
837	1	1	435	gi|1146183	putative [Bacillus subtilis]	72	54	435
868	1	150	788	gi|1377842	unknown [Bacillus subtilis]	72	55	639
922	1	130	432	gi|1088269	unknown protein [Azotobacter vinelandii]	72	58	303
941	1	2	238	gi|153929	NADPH-sulfite reducatase flavoprotein component [Salmonella yphimurium]	72	49	237
980	1	840	421	gi|853767	UDP-N-acetylglucosamine 1-carboxyvinyltransferase [Bacillus ubtilis]	72	59	420

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
1209	1	383	213	gi|144735	neurotoxin type B (Clostridium botulinum)	72	44	171
1469	2	671	474	gi|1205458	hypothetical protein (GB:D26562_47) (Haemophilus influenzae)	72	63	198
1956	1	727	365	gi|154409	hexosephosphate transport protein (Salmonella typhimurium) pir|B41853|B41853 hexose phosphate transport system regulatory rotein uhpB - Salmonella typhimurium	72	44	363
2101	1	3	401	gi|1303950	YqiY (Bacillus subtilis)	72	50	399
2503	1	569	399	gi|149713	formate dehydrogenase (Methanobacterium formicicum) pir|A42712|A42712 formate dehydrogenase (EC 1.2.1.2) - ethanobacterium formicicum	72	56	171
2967	1	3	155	gi|1212729	YqhJ (Bacillus subtilis)	72	46	153
3004	1	367	185	gi|665999	hypothetical protein (Bacillus subtilis)	72	55	183
3109	1	278	141	gi|413968	ipa-44d gene product (Bacillus subtilis)	72	45	138
3171	1	3	287	gi|515938	glutamate synthase (ferredoxin) (Synechocystis sp.) pir|S46957|S46957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - ynechocystis sp.	72	52	285
3771	1	26	367	gi|1408501	homologous to N-acyl-L-amino acid amidohydrolase of Bacillus stearothermophilus (Bacillus subtilis)	72	63	342
3951	1	1	222	gi|1500409	M. jannaschii predicted coding region MJ1519 (Methanococcus jannaschii)	72	38	222
4190	1	721	362	gi|39956	IIGlc (Bacillus subtilis)	72	57	360
4444	1	3	347	gi|1009366	Respiratory nitrate reductase (Bacillus subtilis)	72	55	345
6	2	911	1200	gi|537095	ornithine carbamoyltransferase (Escherichia coli)	71	55	270
11	15	11350	10859	gi|532309	25 kDa protein (Escherichia coli)	71	47	492
19	2	1248	2435	gi|1244574	D-alanine:D-alanine ligase (Enterococcus hirae)	71	52	1188
21	2	898	1488	gi|149629	anthranilate synthase component 2 (Leptospira biflexa) pir|C32840|C32840 anthranilate synthase (EC 4.1.3.27) component II Leptospira biflexa	71	45	591
34	1	1	567	gi|1303983	YqkF (Bacillus subtilis)	71	59	567
37	3	3192	2806	gi|1209681	glutamate-rich protein (Bacillus firmus)	71	50	387
38	18	12250	12462	gi|927645	arginyl endopeptidase (Porphyromonas gingivalis)	71	50	213
39	3	1246	4431	pir|S09411|S094	spoIIIE protein - Bacillus subtilis	71	49	3186
53	14	15770	14760	gi|142611	branched chain alpha-keto acid dehydrogenase E1-alpha (Bacillus ubtilis)	71	58	1011
54	11	13461	12625	gi|143014	gnt repressor (Bacillus subtilis)	71	46	837

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
57	7	7152	5860	gi|508175	EIIC domain of PTS-dependent Gat transport and phosphorylation Escherichia coli	71	48	1293
57	18	13897	14334	gi|1063247	high homology to flavohemoprotein (Haemoglobin-like protein) of Alcaligenes eutrophus and Saccharomyces cerevisiae [Bacillus subtilis]	71	56	438
62	16	9831	10955	gi|1303926	YqiG (Bacillus subtilis)	71	54	1125
70	12	8505	8966	gi|147198	phnE protein (Escherichia coli)	71	38	462
86	5	2394	2089	gi|904205	hypothetical protein (Bacillus subtilis)	71	51	306
96	7	7601	8269	gi|709991	hypothetical protein (Bacillus subtilis)	71	49	669
100	6	4822	5931	gi|1060848	Opine dehydrogenase (Arthrobacter sp.)	71	45	1110
103	1	1062	532	gi|143089	iep protein (Bacillus subtilis)	71	41	531
109	18	15313	15695	gi|413985	ipa-61d gene product (Bacillus subtilis)	71	57	384
113	1	630	316	gi|663254	probable protein kinase (Saccharomyces cerevisiae)	71	57	315
114	5	6598	5603	gi|143156	membrane bound protein (Bacillus subtilis)	71	40	996
133	2	3087	1723	gi|1303913	YqhX (Bacillus subtilis)	71	53	1365
149	19	6335	5895	gi|529650	G40P (Bacteriophage SPP1)	71	51	441
154	5	3635	3087	gi|425488	repressor protein (Streptococcus sobrinus)	71	47	549
164	11	11354	11689	gi|49318	ORF4 gene product (Bacillus subtilis)	71	52	336
169	5	1936	2745	gi|1403403	unknown (Mycobacterium tuberculosis)	71	56	810
193	2	272	1234	gi|1303788	YqeH (Bacillus subtilis)	71	49	963
205	1	1743	895	gi|1215694	GlnQ (Mycoplasma pneumoniae)	71	46	849
233	4	1849	2022	gi|633732	ORF1 (Campylobacter jejuni)	71	50	174
237	7	4501	5169	gi|149384	HisIE (Lactococcus lactis)	71	54	669
272	4	2848	2273	gi|709993	hypothetical protein (Bacillus subtilis)	71	48	576
274	2	618	1496	gi|143035	NAD(P)H:glutamyl-transfer RNA reductase [Bacillus subtilis] pir|A35252|A35252 5-aminolevulinate synthase (EC 2.3.1.37) - acillus subtilis	71	53	879
276	5	3349	2720	gi|303562	ORF210 (Escherichia coli)	71	50	630
287	1	136	660	gi|310634	20 kDa protein (Streptococcus gordonii)	71	53	525
288	6	3322	2771	gi|1256625	putative (Bacillus subtilis)	71	47	552

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
301	6	3492	2461	gi|467417	similar to lysine decarboxylase [Bacillus subtilis]	71	57	1032
306	4	6607	5222	gi|1256618	transport protein [Bacillus subtilis]	71	56	1386
307	2	1536	925	gi|632683	orfC [Mycoplasma capricolum]	71	45	612
310	5	5793	5146	gi|348052	acetoin utilization protein [Bacillus subtilis]	71	51	648
322	1	2	1303	gi|1001819	hypothetical protein [Synechocystis sp.]	71	46	1302
333	4	4171	3995	gi|467473	unknown [Bacillus subtilis]	71	57	177
350	2	548	922	gi|551879	ORF 1 [Lactococcus lactis]	71	55	375
375	4	1860	3071	gi|467447	unknown [Bacillus subtilis]	71	57	1212
380	5	1560	2102	gi|142557	ATP synthase b subunit [Bacillus megaterium]	71	43	543
414	2	251	637	gi|580904	homologous to E.coli rnpA [Bacillus subtilis]	71	49	387
424	1	335	1354	gi|581305	L-lactate dehydrogenase [Lactobacillus plantarum]	71	57	1020
436	4	3701	3270	pir|PN0501|PN05	phosphoribosylanthranilate isomerase (EC 5.3.1.24) - Bacillus subtilis (fragment)	71	66	432
482	1	3	1280	gi|410142	ORFX18 [Bacillus subtilis]	71	49	1278
525	3	2272	1844	gi|143370	phosphoribosylpyrophosphate amidotransferase (PUR-F; EC 2.4.2.14) Bacillus subtilis	71	56	429
529	4	2739	2047	gi|606150	ORF_f309 [Escherichia coli]	71	43	691
563	1	22	969	gi|1237015	ORF4 [Bacillus subtilis]	71	53	948
581	1	506	255	gi|1301730	T25G3.2 [Caenorhabditis elegans]	71	47	252
612	2	1068	913	gi|153968	fimbriae Z [Salmonella typhimurium]	71	55	156
613	1	1	654	gi|466778	lysine specific permease [Escherichia coli]	71	50	654
618	1	1243	623	gi|1146238	poly(A) polymerase [Bacillus subtilis]	71	52	621
630	1	1170	586	gi|1486243	unknown [Bacillus subtilis]	71	53	585
691	1	1126	641	gi|289260	comE ORF1 [Bacillus subtilis]	71	51	486
694	2	149	427	gi|12971	NADH dehydrogenase subunit V (AA 1-605) [Gallus gallus] ir|S10197|S10197 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain - chicken mitochondrion (SGC1)	71	47	279
715	2	169	777	gi|1303830	YqfL [Bacillus subtilis]	71	53	609
746	2	1473	970	gi|1377843	unknown [Bacillus subtilis]	71	52	504

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
748	1	1437	802	gi|1405459	YneS (Bacillus subtilis)	71	49	636
753	1	1018	524	gi|1510389	M. jannaschii predicted coding region MJ0296 (Methanococcus jannaschii)	71	53	495
761	1	3	215	gi|475972	pentafunctional enzyme (Pneumocystis carinii)	71	47	213
783	1	1203	703	gi|536655	ORF YBR244w (Saccharomyces cerevisiae)	71	52	501
800	3	1292	987	gi|1204326	tRNA delta(2)-isopentenylpyrophosphate transferase (Haemophilus influenzae)	71	48	306
806	1	116	286	gi|1419075	cbiH gene product (Methanobacterium thermoautotrophicum)	71	50	171
931	1	973	488	gi|893358	PgsA (Bacillus subtilis)	71	56	486
1041	1	2	262	gi|1408507	pyrimidine nucleoside transport protein (Bacillus subtilis)	71	45	261
1070	1	2	172	gi|709993	hypothetical protein (Bacillus subtilis)	71	46	171
1176	1	57	365	gi|151259	HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas mevalonii) pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	71	49	309
1181	1	366	184	gi|46971	epiP gene product (Staphylococcus epidermidis)	71	50	183
1281	1	3	290	gi|153016	ORF 419 protein (Staphylococcus aureus)	71	50	288
1348	1	456	229	gi|602683	orfC (Mycoplasma capricolum)	71	48	228
2002	1	756	379	gi|1008177	ORF YJL046w (Saccharomyces cerevisiae)	71	48	378
2119	1	2	217	gi|1046088	arginyl-tRNA synthetase (Mycoplasma genitalium)	71	50	216
2418	1	3	320	gi|1499771	M. jannaschii predicted coding region MJ0936 (Methanococcus jannaschii)	71	57	318
2961	1	2	187	gi|312443	carbamoyl-phosphate synthase (glutamine-hydrolysing) (Bacillus aldolyticus)	71	57	186
2999	2	67	306	gi|710020	nitrite reductase (nirB) (Bacillus subtilis)	71	43	240
3033	1	2	184	gi|1262335	YmaA (Bacillus subtilis)	71	57	183
3584	1	3	338	gi|401716	beta-isopropylmalate dehydrogenase (Neurospora crassa)	71	55	336
3715	2	743	399	gi|563952	gluconate permease (Bacillus licheniformis)	71	59	345
3785	1	770	387	gi|47382	acyl-CoA-dehydrogenase (Streptomyces purpurascens)	71	57	384
3875	1	541	272	gi|1001541	hypothetical protein (Synechocystis sp.)	71	38	270
4135	1	637	320	gi|142695	S-adenosyl-L-methionine:uroporphyrinogen III methyltransferase Bacillus megaterium)	71	52	318
4249	1	63	239	gi|1205363	deoxyribose aldolase (Haemophilus influenzae)	71	63	177
4508	1	530	267	gi|1197667	vitellogenin (Anolis pulchellus)	71	46	264

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
6	3	1237	2721	gi|1321788	arginine ornithine antiporter [Clostridium perfringens]	70	54	1485
11	11	6572	7486	gi|216854	P47K [Pseudomonas chlororaphis]	70	41	915
12	1	2890	1481	gi|467330	replicative DNA helicase [Bacillus subtilis]	70	49	1410
15	1	1756	893	gi|451216	Mannosephosphate Isomerase [Streptococcus mutans]	70	46	864
15	2	1277	1050	gi|476092	unknown [Bacillus subtilis]	70	50	228
17	2	2132	1350	gi|145402	choline dehydrogenase [Escherichia coli]	70	52	783
21	1	2	925	gi|149516	anthranilate synthase alpha subunit [Lactococcus lactis] pir|S35124|S35124 anthranilate synthase (EC 4.1.3.27) alpha chain - actococcus lactis subsp. lactis	70	50	924
25	7	5580	6251	gi|1389549	ORF3 [Bacillus subtilis]	70	52	672
33	6	6071	7423	gi|1303875	YqhB [Bacillus subtilis]	70	51	1353
36	2	959	1594	gi|500755	methyl purine glycosylase [Mus musculus]	70	47	636
38	8	4901	5860	gi|1408507	pyrimidine nucleoside transport protein [Bacillus subtilis]	70	44	960
44	8	5312	5989	gi|1006620	hypothetical protein [Synechocystis sp.]	70	49	678
46	10	8950	10020	gi|1403126	czcD gene product [Alcaligenes eutrophus]	70	45	1071
52	2	2727	1900	gi|1486247	unknown [Bacillus subtilis]	70	53	828
52	6	4048	4656	gi|244501	esterase II=carboxylesterase (EC 3.1.1.1) [Pseudomonas fluorescens, eptido, 218 aa]	70	50	609
56	8	8460	9962	gi|1339951	small subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	70	51	1503
62	1	48	290	gi|142702	A competence protein 2 [Bacillus subtilis]	70	47	243
64	1	1080	541	gi|1204377	molybdopterin biosynthesis protein [Haemophilus influenzae]	70	47	540
70	5	5139	3595	gi|1204834	2',3'-cyclic-nucleotide 2'-phosphodiesterase [Haemophilus influenzae]	70	47	1545
91	4	7793	5466	gi|886471	methionine synthase [Catharanthus roseus]	70	56	2328
96	5	8754	7255	pir|B39096|B390	alkaline phosphatase (EC 3.1.3.1) III precursor - Bacillus subtilis	70	54	1500
110	2	767	1300	gi|145294	adenine phosphoribosyl-transferase [Escherichia coli]	70	51	534
116	6	7026	7976	gi|143607	sporulation protein [Bacillus subtilis]	70	50	951
121	8	6401	6988	gi|1107528	ttg start [Campylobacter coli]	70	45	588
131	8	6842	7936	gi|1150454	prolidase PepQ [Lactobacillus delbrueckii]	70	48	1095

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match Acession	match gene name	% sim	% ident	length (nt)
135	1	2	1489	gi|311309	putative membrane-bound protein with four times repitition of ro-Ser-Ala at the N-terminus; function unknown [Alcaligenes utrophus]	70	49	1488
138	3	418	714	gi|904181	hypothetical protein [Bacillus subtilis]	70	46	297
164	8	9344	9874	gi|49315	ORF1 gene product [Bacillus subtilis]	70	47	531
164	16	15626	16618	gi|1205212	hypothetical protein (GB:D10483_18) [Haemophilus influenzae]	70	50	993
205	2	2735	1803	gi|1215695	peptide transport system protein SapF homolog; SapF homolog [Mycoplasma pneumoniae]	70	47	933
209	3	910	1386	gi|1204665	hypothetical protein (GB:X73124_26) [Haemophilus influenzae]	70	48	477
246	3	340	756	gi|215098	excisionase [Bacteriophage 154a]	70	46	417
263	7	7876	6749	gi|142540	aspartokinase II [Bacillus sp.]	70	51	1128
268	3	3212	4117	gi|1340128	ORF1 [Staphylococcus aureus]	70	50	906
302	6	3201	3827	gi|147782	ruvA protein (gtg start) [Escherichia coli]	70	46	627
302	10	5879	7051	pir|C38530|C385	queuine tRNA-ribosyltransferase (EC 2.4.2.29) - Escherichia coli	70	55	1173
313	1	2520	1414	gi|1205934	aminopeptidase a/i [Haemophilus influenzae]	70	46	1107
355	2	379	669	gi|1070013	protein-dependent [Bacillus subtilis]	70	48	291
403	1	1255	629	gi|733147	GumF [Xanthomonas campestris]	70	33	627
444	10	8770	9273	gi|1204752	high affinity ribose transport protein [Haemophilus influenzae]	70	52	504
449	1	2	1243	gi|619724	MgtE [Bacillus firmus]	70	44	1242
472	1	637	320	gi|727145	open reading frame; putative [Bacillus amyloliquefaciens] pir|B29091|B29091 hypothetical protein (bglA region) - Bacillus myloliquefaciens (fragment)	70	41	318
480	2	727	1608	gi|142560	ATP synthase gamma subunit [Bacillus megaterium]	70	44	882
524	1	2	307	gi|602292	RCH2 protein [Brassica napus]	70	45	306
525	1	823	413	gi|143372	phosphoribosyl glycinamide formyltransferase (PUR-N) [Bacillus ubtilis]	70	52	411
565	4	3625	2552	gi|881434	ORFP [Bacillus subtilis]	70	51	1074
607	4	829	1284	gi |511524	hypothetical protein (SP:P37002) [Methanococcus jannaschii]	70	50	456
633	1	1383	703	gi|431231	uracil permease [Bacillus caldolyticus]	70	53	681
646	3	1683	1309	gi|467340	unknown [Bacillus subtilis]	70	49	375
663	1	830	417	gi|1303873	YqjZ [Bacillus subtilis]	70	40	414

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
681	1	1488	781	gi|1001678	hypothetical protein [Synechocystis sp.]	70	53	708
708	1	2	448	sp|P33940|YOJH_	HYPOTHETICAL 54.3 KD PROTEIN IN ECO-ALKB INTERGENIC REGION.	70	51	447
725	1	51	722	gi|1001644	hypothetical protein [Synechocystis sp.]	70	48	672
776	1	1371	787	gi|145165	putative [Escherichia coli]	70	47	585
834	1	250	783	gi|552971	NADH dehydrogenase (ndhF) [Vicia faba]	70	47	534
865	2	1585	1379	gi|1204636	ATP-dependent helicase [Haemophilus influenzae]	70	45	207
894	1	535	269	gi|467364	DNA binding protein (probale) [Bacillus subtilis]	70	41	267
919	1	3	317	gi|1314847	CinA [Bacillus subtilis]	70	40	315
944	1	3	572	gi|709991	hypothetical protein [Bacillus subtilis]	70	44	570
988	2	772	605	gi|142441	ORF 3; putative [Bacillus subtilis]	70	50	168
1055	1	3	335	gi|529755	speC [Streptococcus pyogenes]	70	37	333
1093	1	2	904	gi|853754	ABC transporter [Bacillus subtilis]	70	49	903
1109	1	2	310	gi|1001827	hypothetical protein [Synechocystis sp.]	70	42	309
1220	1	468	235	pir|S23416|S234	epiB protein - Staphylococcus epidermidis	70	40	234
1279	1	73	348	gi|153015	FemA protein [Staphylococcus aureus]	70	47	276
1330	1	195	542	sp|P31776|PBPA_	PENICILLIN-BINDING PROTEIN 1A (PBP-1A) (PENICILLIN-BINDING PROTEIN A).	70	50	348
1537	2	232	402	gi|1146181	putative [Bacillus subtilis]	70	50	171
1574	1	451	272	gi|219630	endothelin-A receptor [Homo sapiens]	70	47	180
1640	1	690	346	gi|1146243	22.4% identity with Escherichia coli DNA-damage inducible protein ...; putative [Bacillus subtilis]	70	46	345
2504	1	2	286	gi|495179	transmembrane protein [Lactococcus lactis]	70	51	285
3061	1	564	301	gi|508175	EIIC domain of PTS-dependent Gat transport and phosphorylation Escherichia coli]	70	44	264
3128	1	2	199	gi|1340096	unknown [Mycobacterium tuberculosis]	70	51	198
3218	1	3	488	gi|515938	glutamate synthase (ferredoxin) [Synechocystis sp.] pir|S46957|S46957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - ynechocystis sp.	70	50	486
3323	1	794	399	gi|1154891	ATP binding protein [Phormidium laminosum]	70	52	396
3679	1	599	399	gi|529385	chromosome condensation protein [Caenorhabditis elegans]	70	30	201

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
3841	1	706	398	gi|1208965	hypothetical 23.3 kd protein (Escherichia coli)	70	47	309
3929	1	3	401	gi|149435	putative (Lactococcus lactis)	70	49	399
4044	1	595	374	gi|602031	similar to trimethylamine DH (Mycoplasma capricolum) pir|S49950|S49950 probable trimethylamine dehydrogenase (EC .5.99.7) - Mycoplasma capricolum (SGC3) (fragment)	70	40	222
4329	1	558	280	gi|1339951	small subunit of NADH-dependent glutamate synthase (Plectonema boryanum)	70	49	279
4422	1	576	289	gi|296464	ATPase (Lactococcus lactis)	70	57	288
4647	1	361	200	gi|166412	NADH-glutamate synthase (Medicago sativa)	70	59	162
16	8	7571	9031	gi|1499620	M. jannaschii predicted coding region MJ0798 (Methanococcus jannaschii)	69	44	1461
16	9	9080	10033	gi|1353197	thioredoxin reductase (Eubacterium acidaminophilum)	69	54	954
30	1	1452	727	gi|1204910	hypothetical protein (GB:U14003_302) (Haemophilus influenzae)	69	52	726
38	4	1023	1298	gi|407773	devA gene product (Anabaena sp.)	69	41	276
44	9	5987	6595	gi|1205920	molybdate uptake system hydrophilic membrane-bound protein (Haemophilus influenzae)	69	45	609
62	15	9104	9475	gi|385178	unknown (Bacillus subtilis)	69	44	372
66	4	2402	2803	gi|1303893	YqhL (Bacillus subtilis)	69	51	402
67	15	14124	13627	gi|149647	ORF2 (Listeria monocytogenes)	69	37	498
67	17	14053	14382	gi|305002	ORF_f356 (Escherichia coli)	69	49	330
67	19	15130	15807	gi|1109684	ProV (Bacillus subtilis)	69	45	678
78	3	1447	2124	gi|1256633	putative (Bacillus subtilis)	69	53	678
78	4	4513	3725	gi|1103958	YqjG (Bacillus subtilis)	69	32	789
85	4	4521	4213	pir|E29326|E293	hypothetical protein (pur operon) - Bacillus subtilis	69	32	309
86	6	3253	2654	gi|973332	OrfC (Bacillus subtilis)	69	50	600
95	1	96	710	gi|78646R	4A11 antigen, sperm tail membrane antigen=putative sucrose-specific phosphotransferase enzyme II homolog (mice, testis, Peptide Partial, 72 aa)	69	43	615
100	7	6023	7426	gi|1205355	Na+/H+ antiporter (Haemophilus influenzae)	69	39	1404
102	2	2678	1650	gi|561690	sialoglycoprotease (Pasteurella haemolytica)	69	47	1029
103	8	12241	8537	gi|1009366	Respiratory nitrate reductase (Bacillus subtilis)	69	54	3705
103	11	14987	12552	gi|710020	nitrite reductase (nirB) (Bacillus subtilis)	69	51	2436

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
112	11	8708	10168	gi|154411	hexosephosphate transport protein (Salmonella typhimurium) pir|D41853|D41853 hexose phosphate transport system protein uhpT - almonella typhimurium	69	51	1461
112	16	16644	17414	gi|1204435	pyruvate formate-lyase activating enzyme (Haemophilus influenzae)	69	50	771
113	2	33	953	gi|290509	o307 (Escherichia coli)	69	43	921
114	2	1537	1058	pir|A42771|A427	reticulocyte-binding protein 1 - Plasmodium vivax	69	39	480
121	6	4309	5310	gi|1154633	NrdF (Bacillus subtilis)	69	53	1002
125	2	267	854	gi|413931	ipa-7d gene product (Bacillus subtilis)	69	43	588
149	27	10666	10400	pir|S28089|S280	hypothetical protein A - yeast (Zygosaccharomyces bisporus) plasmid pSB)	69	39	267
161	1	1598	813	gi|1205538	hypothetical protein (GB:U14003_302) (Haemophilus influenzae)	69	47	786
165	4	2222	4633	gi|40054	phenylalanyl-tRNA synthetase beta subunit (AA 1-804) (Bacillus btilis)	69	52	2412
169	3	1210	1761	gi|296031	elongation factor Ts (Spirulina platensis)	69	45	552
175	12	8686	8339	gi|732682	FimE protein (Escherichia coli)	69	69	348
190	2	484	1671	sp|P17731|HIS8_	HISTIDINOL-PHOSPHATE AMINOTRANSFERASE (EC 2.6.1.9) (IMIDAZOLE ACETOL-PHOSPHATE TRANSAMINASE).	69	48	1188
206	1	5551	2777	gi|41750	hsdR protein (AA 1-1033) (Escherichia coli)	69	49	2775
206	4	6038	5796	gi|1256135	YbbF (Bacillus subtilis)	69	48	243
249	1	636	319	gi|1405456	YmuP (Bacillus subtilis)	69	50	318
302	8	4820	5776	gi|1001768	hypothetical protein (Synechocystis sp.)	69	48	957
324	2	7384	3893	gi|1256798	pyruvate carboxylase (Rhizobium etli)	69	53	3492
351	3	2098	1808	gi|1491664	T04H1.4 (Caenorhabditis elegans)	69	30	291
369	3	2075	2305	gi|336458	ORF (Balaenoptera acutorostrata)	69	61	231
392	3	1999	2424	gi|556015	ORF1 (Bacillus subtilis)	69	45	426
410	1	87	779	gi|155611	phosphoglyceromutase (Zymomonas mobilis)	69	58	693
421	1	2085	1129	gi|1276985	arginase (Bacillus caldovelox)	69	54	957
444	8	6713	7741	gi|1221782	purine synthesis repressor (Haemophilus influenzae)	69	40	1029
453	1	828	415	gi|1122758	unknown (Bacillus subtilis)	69	57	414
469	2	3286	2246	gi|1458228	mutY homolog (Homo sapiens)	69	44	1041

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
509	3	1730	1371	gi|49224	URF 4 [Synechococcus sp.]	69	39	360
520	5	3023	2823	gi|726427	similar to D. melanogaster MST101-2 protein (PIR:S34154) Caenorhabditis elegans]	69	39	201
531	1	26	760	gi|509672	repressor protein [Bacteriophage Tuc2009]	69	33	735
589	1	107	253	gi|169101	17.9 kDa heat shock protein (hsp17.9) [Pisum sativum]	69	52	147
594	2	597	1391	gi|142783	DNA photolyase [Bacillus firmus]	69	48	795
604	4	2476	2114	gi|413930	ipa-6d gene product [Bacillus subtilis]	69	45	363
607	1	2	313	gi|1236103	W08D2.3 [Caenorhabditis elegans]	69	47	312
607	2	590	312	gi|536715	ORF YBR275c [Saccharomyces cerevisiae]	69	39	279
734	1	864	433	gi|467327	unknown [Bacillus subtilis]	69	44	432
759	1	3	338	gi|1009367	Respiratory nitrate reductase [Bacillus subtilis]	69	50	336
761	2	392	586	gi|3508	Leucyl-tRNA synthetase (cytoplasmic) [Saccharomyces cerevisiae] |1370340 ORF YPL160w [Saccharomyces cerevisiae]	69	46	195
802	1	72	1013	gi|143044	ferrochelatase [Bacillus subtilis]	69	55	942
816	1	2573	1368	gi|1510268	restriction modification system S subunit [Methanococcus jannaschii]	69	45	1206
838	2	133	387	gi|1255371	coded for by C. elegans cDNA yk34a9.5; coded for by C. elegans cDNA yk34a9.3; Similar to guanylate kinase [Caenorhabditis elegans]	69	46	255
851	2	745	1005	gi|288998	secA gene product [Antithamnion sp.]	69	39	261
867	1	535	269	gi|1070014	protein-dependent [Bacillus subtilis]	69	47	267
995	1	954	478	gi|1205569	transcription elongation factor [Haemophilus influenzae]	69	53	477
999	1	1009	506	gi|899256	predicted trithorax protein [Drosophila virilis]	69	21	504
1127	1	1315	659	gi|1205434	H. influenzae predicted coding region HI1191 [Haemophilus influenzae]	69	56	657
1138	1	248	460	gi|1510646	H. jannaschii predicted coding region MJ0568 [Methanococcus jannaschii]	69	48	213
2928	1	3	401	gi|290503	glutamate permease [Escherichia coli]	69	41	399
3090	1	444	223	gi|1204987	DNA polymerase III, alpha chain [Haemophilus influenzae]	69	36	222
3817	1	2	400	gi|1483199	peptide-synthetase [Amycolatopsis mediterranei]	69	45	399
3833	1	667	335	gi|1524193	unknown [Mycobacterium tuberculosis]	69	46	333

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
4079	1	747	400	gi|546918	orfY 3' of comK (Bacillus subtilis, E26, Peptide Partial, 140 aa) pir|S43612|S43612 hypothetical protein Y - Bacillus subtilis sp|P40398|YHXD_BACSU HYPOTHETICAL PROTEIN IN COMK 3'REGION (ORFY) FRAGMENT).	69	64	348
4115	2	215	400	gi|517205	67 kDa Myosin-crossreactive streptococcal antigen (Streptococcus yogenes)	69	59	186
4139	1	1	333	gi|1208451	hypothetical protein (Synechocystis sp.)	69	36	333
4258	1	457	230	gi|496158	restriction-modification enzyme subunit M1 (Mycoplasma pulmonis) pir|S49395|S49395 HsdM1 protein - Mycoplasma pulmonis (SGC3)	69	43	228
4317	1	90	374	gi|413967	ipa-43d gene product (Bacillus subtilis)	69	44	285
4465	1	3	293	gi|396296	similar to phosphotransferase system enzyme II (Escherichia coli) sp|P32672|PTWC_ECOLI PTS SYSTEM, FRUCTOSE-LIKE-2 IIC COMPONENT PHOSPHOTRANSFERASE ENZYME II, C COMPONENT).	69	49	291
3	1	2302	1193	gi|1109685	ProW (Bacillus subtilis)	68	46	1110
15	4	2592	2074	gi|807973	unknown (Saccharomyces cerevisiae)	68	45	519
31	8	6328	8772	gi|290642	ATPase (Enterococcus hirae)	68	48	2445
40	2	1115	750	gi|606342	ORF_o622; reading frame open far upstream of start; possible rameshift, linking to previous ORF (Escherichia coli)	68	55	166
46	9	6886	8415	gi|155276	aldehyde dehydrogenase (Vibrio cholerae)	68	44	1530
48	3	3643	3404	gi|285608	241k polyprotein (Apple stem grooving virus)	68	47	240
48	4	3536	4132	gi|1045937	M. genitalium predicted coding region MG246 (Mycoplasma genitalium)	68	39	597
53	10	11671	10685	gi|1303952	YqjA (Bacillus subtilis)	68	46	987
70	9	7346	8155	gi|147198	phnE protein (Escherichia coli)	68	40	810
89	4	1899	2966	gi|145173	35 kDa protein (Escherichia coli)	68	43	1068
108	1	2187	1150	gi|38722	precursor (aa -20 to 381) (Acinetobacter calcoaceticus) ir|A29277|A29277 aldose 1-epimerase (EC 5.1.3.3) - Acinetobacter lcoaceticus	68	57	1038
112	5	2666	3622	gi|153724	HalC (Streptococcus pneumoniae)	68	55	957
116	7	7865	8638	gi|143608	sporulation protein (Bacillus subtilis)	68	48	774
118	3	2484	3698	gi|1303805	YqeR (Bacillus subtilis)	68	46	1215
120	2	1424	1594	sp|P38038|CYSJ_	SULFITE REDUCTASE (NADPH) FLAVOPROTEIN ALPHA-COMPONENT (EC 1.8.1.2) (SIR-FP).	68	45	171
129	1	1	1011	gi|396307	argininosuccinate lyase (Escherichia coli)	68	50	1011

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
132	3	1867	2739	gi|216267	ORF2 (Bacillus megaterium)	68	48	873
134	2	848	1012	gi|147545	DNA recombinase (Escherichia coli)	68	50	165
141	2	372	614	gi|872116	sti (stress inducible protein) (Glycine max)	68	36	243
149	7	2454	2260	gi|145774	hsp70 protein (dnaK gene) (Escherichia coli)	68	48	195
155	2	1776	1534	gi|216583	ORF1 (Escherichie coli)	68	38	243
158	3	1826	3289	sp|P33940|YOJH_	HYPOTHETICAL 54.3 KD PROTEIN IN ECO-ALKB INTERGENIC REGION.	68	51	1464
169	6	2749	3318	gi|1403402	unknown (Mycobacterium tuberculosis)	68	46	570
175	10	9158	7365	gi|1072395	phaA gene product (Rhizobium meliloti)	68	51	1794
188	7	4184	5434	gi|1173843	3-ketoacyl-ACP synthase II (Vibrio harveyi)	68	48	1251
189	3	907	1665	gi|467383	DNA binding protein (probable) (Bacillus subtilis)	68	55	759
206	5	7683	6709	gi|1256138	YbbI (Bacillus subtilis)	68	48	975
206	8	10425	12176	gi|452687	pyruvate decarboxylase (Saccharomyces cerevisiae)	68	48	1752
212	8	3421	3648	gi|1369941	c1 gene product (Bacteriophage B1)	68	39	228
214	8	5457	6482	gi|1420467	ORF YOR196c (Saccharomyces cerevisiae)	68	45	1026
237	4	2507	3088	gi|149381	HisH (Lactococcus lactis)	68	46	582
243	5	5540	4542	gi|1235684	mevalonate pyrophosphate decarboxylase (Saccharomyces cerevisiae)	68	47	999
262	1	3	164	gi|150974	4-oxalocrotonate tautomerase (Pseudomonas putida)	68	42	162
262	2	1984	1118	gi|1147744	PSR (Enterococcus hirae)	68	49	867
276	6	3702	3139	sp|P30750|ABC_E	ATP-BINDING PROTEIN ABC (FRAGMENT).	68	50	564
306	6	6345	5725	gi|1256617	adenine phosphoribosyltransferase (Bacillus subtilis)	68	53	621
333	3	4599	3850	gi|467473	unknown (Bacillus subtilis)	68	45	750
365	6	5017	4838	gi|1130643	T22B3.3 (Caenorhabditis elegans)	68	45	180
376	2	549	1646	gi|1277026	DAPA aminotransferase (Bacillus subtilis)	68	51	1098
405	1	1741	872	gi|1303917	YqiB (Bacillus subtilis)	68	47	870
406	2	853	539	gi|1511513	ABC transporter, probable ATP-binding subunit (Methanococcus jannaschii)	68	44	315
426	6	3558	3391	gi|624632	GltL (Escherichia coli)	68	48	168
438	1	108	329	gi|146923	nitrogenase reductase (Escherichia coli)	68	43	222

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
443	1	476	240	gi|535810	hippuricase (Campylobacter jejuni)	68	42	237
443	2	518	1015	gi|1204742	H. influenzae predicted coding region HI0491 (Haemophilus influenzae)	68	48	498
443	5	4447	3779	gi|809660	deoxyribose-phosphate aldolase (Bacillus subtilis) pir|S49455|S49455 deoxyribose-phosphate aldolase (EC 4.1.2.4) - acillus subtilis	68	55	669
476	2	240	1184	gi|571345	unknown, similar to E.coli cardiolipin synthase (Bacillus subtilis) sp|P45860|YWIE_BACSU HYPOTHETICAL 58.2 PROTEIN IN NARI-ACDA NTERGENIC REGION.	68	45	945
486	2	1876	1046	gi|147328	transport protein (Escherichia coli)	68	41	831
517	3	1764	2084	gi|523809	orf2 (Bacteriophage A2)	68	64	321
572	1	2	571	sp|P39237|YO5L_	HYPOTHETICAL 6.8 KD PROTEIN IN NRDC-TK INTERGENIC REGION.	68	47	570
646	1	914	459	gi|413982	ipa-58r gene product (Bacillus subtilis)	68	52	456
659	3	1668	1901	gi|1107541	C33D9.8 (Caenorhabditis elegans)	68	36	234
864	5	1510	1716	gi|145774	hsp70 protein (dnaK gene) (Escherichia coli)	68	48	207
920	1	860	432	gi|1510416	hypothetical protein (SP:P31466) (Methanococcus jannaschii)	68	54	429
952	1	1096	611	gi|603456	reductase (Leishmania major)	68	46	486
970	1	91	402	gi|1354775	pfoS/R (Treponema pallidum)	68	46	312
1028	1	1064	534	gi|490117	diaminopimelate decarboxylase (Bacillus subtilis)	68	47	531
1039	1	428	216	gi|1335714	Plasmodium falciparum mRNA for asparagine-rich antigen (clone 17C1) (Plasmodium falciparum)	68	31	213
1058	1	692	348	gi|581649	epiC gene product (Staphylococcus epidermidis)	68	46	345
1096	2	665	465	gi|143434	Rho Factor (Bacillus subtilis)	68	43	201
1308	1	2	694	gi|169939	group B oligopeptidase, PepB (Streptococcus agalactiae)	68	50	693
1679	1	2	238	gi|517205	67 kDa Myosin-crossreactive streptococcal antigen (Streptococcus yogenes)	68	53	237
2039	1	3	383	gi|153898	transport protein (Salmonella typhimurium)	68	51	381
2077	1	3	326	pir|C33496|C334	hisC homolog - Bacillus subtilis	68	47	324
2112	1	613	374	gi|64884	lamin LII (Xenopus laevis)	68	50	240
2273	1	793	398	gi|581648	epiB gene product (Staphylococcus epidermidis)	68	45	396
2948	1	2	385	gi|216869	branched-chain amino acid transport carrier (Pseudomonas aeruginosa) pir|A38534|A38534 branched-chain amino acid transport protein braZ Pseudomonas aeruginosa	68	41	384

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
2955	1	768	400	gi|904179	hypothetical protein [Bacillus subtilis]	68	49	369
2981	1	572	288	gi|508979	GTP-binding protein [Bacillus subtilis]	68	48	285
3014	1	584	294	gi|1524394	ORF-2 upstream of gbsAB operon [Bacillus subtilis]	68	45	291
3082	1	336	169	gi|1204696	fructose-permease IIBC component [Haemophilus influenzae]	68	53	168
3108	1	103	258	gi|217855	heat-shock protein [Arabidopsis thaliana]	68	48	156
3639	1	919	461	gi|1510490	nitrate transport permease protein [Methanococcus jannaschii]	68	47	459
3657	1	1	330	gi|155369	PTS enzyme-II fructose [Xanthomonas campestris]	68	48	330
3823	1	780	391	gi|603768	HutI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis] gi|603768 HutI protein, imidazolone-5-propionate hydrolase Bacillus subtilis]	68	54	390
3982	1	2	277	gi|149435	putative [Lactococcus lactis]	68	47	276
4051	1	1	342	gi|450688	hsdM gene of EcoprrI gene product [Escherichia coli] pir|S38437|S38437 hsdM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	68	48	342
4089	1	12	209	gi|1353678	heavy-metal transporting P-type ATPase [Proteus mirabilis]	68	47	198
4143	1	47	187	gi|603769	HutU protein, urocanase [Bacillus subtilis]	68	55	141
4148	1	2	352	gi|450688	hsdM gene of EcoprrI gene product [Escherichia coli] pir|S38437|S38437 hsdM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	68	51	351
4173	1	2	382	gi|1041097	Pyruvate Kinase [Bacillus psychrophilus]	68	48	381
4182	1	498	250	gi|413968	ipa-44d gene product [Bacillus subtilis]	68	50	249
4362	2	148	318	gi|450688	hsdM gene of EcoprrI gene product [Escherichia coli] pir|S38437|S38437 hsdM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	68	44	171
5	11	9493	8300	gi|143727	putative [Bacillus subtilis]	67	46	1194
31	11	10318	9833	gi|216746	D-lactate dehydrogenase [Lactobacillus plantarum]	67	41	486
32	3	3560	3155	gi|1098557	renal sodium/dicarboxylate cotransporter [Homo sapiens]	67	46	1596
32	5	4945	4145	gi|1510720	prephenate dehydratase [Methanococcus jannaschii]	67	51	801
36	5	5350	4268	gi|[illegible]46216	45% identity with the product of the ORF6 gene from the Erwinia herbicola carotenoid biosynthesis cluster; putative [Bacillus subtilis]	67	58	1083
44	7	4492	5304	gi|1006621	hypothetical protein [Synechocystis sp.]	67	43	813

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
56	7	3943	8481	gi|304131	glutamate synthase large subunit precursor [Azospirillum brasilense] pir|B46602|B46602 glutamate synthase (NADPH) (EC 1.4.1.13) alpha hain - Azospirillum brasilense	67	52	4539
56	12	13923	14678	gi|1000453	TroR [Bacillus subtilis]	67	48	756
62	8	5092	4757	gi|1113949	orf3 [Bacillus, C-125, alkali-sensitive mutant 18224, Peptide Mutant, 112 aa]	67	45	336
62	10	7570	6338	gi|654655	Na/H antiporter system [Bacillus alcalophilus]	67	49	1233
99	3	2119	3321	gi|1204349	hypothetical protein (GB:GB:D90212_3) [Haemophilus influenzae]	67	50	1203
102	9	5695	7176	gi|149432	putative [Lactococcus lactis]	67	51	1482
103	13	14549	14049	gi|1408497	LP9D gene product [Bacillus subtilis]	67	48	501
109	15	14821	13982	gi|413976	ipa-52r gene product [Bacillus subtilis]	67	49	840
109	17	14811	15194	gi|413983	ipa-59d gene product [Bacillus subtilis]	67	29	384
121	4	1713	2153	gi|1262335	YmaA [Bacillus subtilis]	67	54	441
122	1	1	1149	gi|143067	ORFB [Bacillus subtilis]	67	35	1149
124	5	4060	3518	gi|556885	Unknown [Bacillus subtilis]	67	47	543
131	2	4584	3589	gi|1046081	hypothetical protein (GB:D26185_10) [Mycoplasma genitalium]	67	30	996
140	3	2899	2297	gi|146549	kdpC [Escherichia coli]	67	45	603
142	4	5409	4198	gi|1212775	GTP cyclohydrolase II [Bacillus amyloliquefaciens]	67	55	1212
147	5	2913	2374	gi|1303709	YrkJ [Bacillus subtilis]	67	44	540
152	8	6341	6673	gi|1377841	unknown [Bacillus subtilis]	67	48	333
161	4	2720	3763	gi|496319	SphX [Synechococcus sp.]	67	47	1044
163	6	1989	3428	gi|595681	2-oxoglutarate/malate translocator [Spinacia oleracea]	67	47	1440
193	3	1351	1626	gi|1511101	shikimate 5-dehydrogenase [Methanococcus jannaschii]	67	53	276
200	2	917	2179	gi|142439	ATP-dependent nuclease [Bacillus subtilis]	67	48	1263
206	10	12445	12801	sp|P37347|YECD_	HYPOTHETICAL 21.8 KD PROTEIN IN ASPS 5'REGION.	67	47	357
206	11	13047	14432	gi|732813	branched-chain amino acid carrier [Lactobacillus delbrueckii]	67	46	1386
208	2	1321	809	gi|1033037	100 kDa heat shock protein (Hsp100) [Leishmania major]	67	36	513
238	3	1039	2052	gi|809542	CbrB protein [Erwinia chrysanthemi]	67	42	1014

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
246	2	176	367	gi|215098	excisionase [Bacteriophage 154a]	67	37	192
276	2	2260	1412	gi|303560	ORF271 [Escherichia coli]	67	50	849
297	6	2223	3056	gi|142784	CtaA protein [Bacillus firmus]	67	46	834
307	7	5220	4186	gi|1070013	protein-dependent [Bacillus subtilis]	67	43	1035
316	1	36	1028	gi|1161061	dioxygenase [Methylobacterium extorquens]	67	52	993
324	3	5650	5030	gi|1469784	putative cell division protein ftsW [Enterococcus hirae]	67	49	621
336	1	524	264	gi|173122	urea amidolyase [Saccharomyces cerevisiae]	67	45	261
360	1	108	1394	sp|P30053|SYH_S	HISTIDYL-TRNA SYNTHETASE (EC 6.1.1.21) (HISTIDINE--TRNA LIGASE) (HISRS).	67	47	1287
364	3	4890	3592	gi|151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mevalonii] pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	67	46	1299
365	3	2940	2113	gi|1296823	orf2 gene product [Lactobacillus helveticus]	67	47	828
367	2	325	918	gi|1039479	ORFU [Lactococcus lactis]	67	47	594
395	3	666	1271	gi|1204516	hypothetical protein (GB:U00014_4) [Haemophilus influenzae]	67	55	606
415	1	1800	901	gi|882579	CG Site No. 29739 [Escherichia coli]	67	46	900
419	1	1799	903	gi|520752	putative [Bacillus subtilis]	67	48	897
474	1	2	796	gi|886906	argininosuccinate synthetase [Streptomyces clavuligerus] pir|S57659|S57659 argininosuccinate synthase (EC 6.3.4.5) - streptomyces clavuligerus	67	49	795
485	2	1921	2226	gi|143434	Rho Factor [Bacillus subtilis]	67	43	306
596	1	1728	865	gi|1303853	YqgF [Bacillus subtilis]	67	47	864
700	1	433	218	gi|1204628	hypothetical protein (SP:P21498) [Haemophilus influenzae]	67	47	216
806	2	249	647	gi|677947	AppC [Bacillus subtilis]	67	51	399
828	2	340	900	gi|777761	IrrA [Synechococcus sp.]	67	37	561
833	1	1407	916	gi|142996	regulatory protein [Bacillus subtilis]	67	41	492
856	1	1555	779	gi|780224	ZK970.2 [Caenorhabditis elegans]	67	38	777
888	1	1614	850	gi|437315	TTG start codon [Bacillus licheniformis]	67	40	765
1034	1	1190	597	gi|1205113	hypothetical protein (GB:L19201_15) [Haemophilus influenzae]	67	45	594
1062	1	636	319	gi|1303850	YqgC [Bacillus subtilis]	67	41	318
1067	1	918	460	pir|A32950|A329	probable reductase protein - Leishmania major	67	54	459

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
1358	1	3	293	gi|1001369	hypothetical protein (Synechocystis sp.)	67	44	291
2181	1	3	302	gi|1510416	hypothetical protein (SP:P31466) (Methanococcus jannaschii)	67	48	300
3000	1	1	507	gi|517205	67 kDa Myosin-crossreactive streptococcal antigen (Streptococcus yogenes)	67	56	507
3066	1	464	234	gi|308861	GTG start codon (Lactococcus lactis)	67	46	231
3087	1	454	251	gi|1205366	oligopeptide transport ATP-binding protein (Haemophilus influenzae)	67	44	204
3101	1	2	256	gi|1531541	uroporphyrinogen III methyltransferase (Zea mays)	67	55	255
3598	1	728	393	gi|151259	HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas mevalonii) pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	67	56	336
3765	2	584	366	gi|557489	menD (Bacillus subtilis)	67	45	219
3788	1	658	398	pir|S52915|S529	nitrate reductase alpha chain - Bacillus subtilis (fragment)	67	45	261
3883	1	2	265	gi|704397	cystathionine beta-lyase (Arabidopsis thaliana)	67	46	264
3926	1	2	340	gi|1483199	peptide-synthetase (Amycolatopsis mediterranei)	67	44	339
4417	1	82	396	gi|1205337	ribonucleotide transport ATP-binding protein (Haemophilus influenzae)	67	46	315
2	3	3075	3989	gi|535348	CodV (Bacillus subtilis)	66	42	915
15	6	2273	2542	gi|46491	SmtB (Synechococcus PCC7942)	66	37	270
31	9	8059	7826	gi|292046	mucin (*Homo sapiens*)	66	44	234
31	10	9034	9258	gi|1204545	mercury scavenger protein (Haemophilus influenzae)	66	48	225
32	6	6347	5253	gi|998342	inducible nitric oxide synthase (Gallus gallus)	66	47	1095
44	13	8856	10124	gi|1510751	molybdenum cofactor biosynthesis moeA protein (Methanococcus jannaschii)	66	46	1269
48	2	1276	2868	gi|150209	ORF 1 (*Mycoplasma mycoides*)	66	40	1593
58	8	7178	8428	gi|665999	hypothetical protein (Bacillus subtilis)	66	47	1251
62	7	5143	4370	gi|1072398	phaD gene product (Rhizobium meliloti)	66	40	774
70	14	11693	10998	gi|809660	deoxyribose-phosphate aldolase (Bacillus subtilis) pir|S49455|S49455 deoxyribose-phosphate aldolase (EC 4.1.2.4) - acillus subtilis	66	55	696
76	1	1	1305	gi|142440	ATP-dependent nuclease (Bacillus subtilis)	66	42	1305
91	6	9236	8205	gi|704397	cystathionine beta-lyase (Arabidopsis thaliana)	66	43	1032
102	5	3810	3265	gi|1204323	hypothetical protein (SP:P31805) (Haemophilus influenzae)	66	41	546

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
103	4	3418	2732	gi|971344	nitrate reductase gamma subunit (Bacillus subtilis) sp|P42177|NARI_BACSU NITRATE REDUCTASE GAMMA CHAIN (EC 1.7.99.4). gi|1009369 Respiratory nitrate reductase (Bacillus subtilis) (SUB -160)	66	48	687
109	6	4243	4674	gi|170886	glucosamine-6-phosphate deaminase (Candida albicans) pir|A46652|A46652 glucosamine-6-phosphate isomerase (EC 5.3.1.10) - east (Candida albicans)	66	45	432
112	17	17491	17712	gi|1323179	ORF YGR111w (Saccharomyces cerevisiae)	66	33	222
116	2	4667	2637	gi|1491813	gamma-glutamyltranspeptidase (Bacillus subtilis)	66	43	2031
150	5	3189	2989	gi|1146224	putative (Bacillus subtilis)	66	30	201
172	5	3264	3662	gi|755152	highly hydrophobic integral membrane protein (Bacillus subtilis) sp|P42953|TAGG_BACSU TEICHOIC ACID TRANSLOCATION PERMEASE PROTEIN AGG.	66	41	399
174	5	4592	3723	gi|1146241	pantothenate synthetase (Bacillus subtilis)	66	49	870
175	4	3209	2880	gi|642655	unknown (Rhizobium meliloti)	66	29	330
175	11	8743	7994	gi|854655	Na/H antiporter system (Bacillus alcalophilus)	66	43	750
190	5	7079	5727	gi|451072	di-tripeptide transporter (Lactococcus lactis)	66	40	1353
195	15	13919	13713	gi|1322411	unknown (Mycobacterium tuberculosis)	66	42	207
217	3	2822	2595	gi|1143542	alternative stop codon (Rattus norvegicus)	66	36	228
233	9	7133	6135	gi|1458327	F08F3.4 gene product (Caenorhabditis elegans)	66	47	999
238	1	41	1041	gi|809541	CbrA protein (Erwinia chrysanthemi)	66	42	999
241	1	2102	1053	gi|153067	peptidoglycan hydrolase (Staphylococcus aureus)	66	53	1050
261	1	1178	648	gi|1510859	M. jannaschii predicted coding region MJ0790 (Methanococcus jannaschii)	66	40	531
263	3	3731	2973	gi|1205865	tetrahydrodipicolinate N-succinyltransferase (Haemophilus influenzae)	66	47	759
272	8	6548	5484	gi|882101	high affinity nickel transporter (Alcaligenes eutrophus) sp|P23516|HOXN_ALCEU HIGH-AFFINITY NICKEL TRANSPORT PROTEIN.	66	44	1065
276	3	2805	2104	gi|1208965	hypothetical 23.3 kd protein (Escherichia coli)	66	47	702
278	2	2830	1784	gi|1488662	phosphatase-associated protein (Bacillus subtilis)	66	48	1047
278	3	3830	2952	gi|303560	ORF271 (Escherichia coli)	66	45	879
279	2	3894	2218	gi|1185289	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1- carboxylate synthase (Bacillus subtilis)	66	48	1677
288	4	2535	2275	gi|1256625	putative (Bacillus subtilis)	66	42	261
292	2	1133	942	gi|1511604	M. jannaschii predicted coding region MJ1651 (Methanococcus jannaschii)	66	30	192

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
294	1	1116	559	gi|216314	esterase [Bacillus stearothermophilus]	66	45	558
297	4	2913	1978	gi|994794	cytochrome a assembly facto [Bacillus subtilis] sp|P24009|COXX_BACSU PROBABLE CYTOCHROME C OXIDASE ASSEMBLY FACTOR.	66	45	936
316	4	2053	2682	gi|1107839	alginate lyase [Pseudomonas aeruginosa]	66	40	630
338	4	2460	2302	gi|520750	biotin synthetase [Bacillus sphaericus]	66	58	159
339	1	1214	735	gi|467468	7, 8-dihydro-6-hydroxymethylpterin-pyrophosphokinase [Bacillus ubtilis]	66	52	480
363	1	3	863	gi|581649	epiC gene product [Staphylococcus epidermidis]	66	47	861
366	2	232	483	gi|1103505	unknown [Schizosaccharomyces pombe]	66	53	252
367	4	2468	1845	sp|P20692|TYRA_	PREPHENATE DEHYDROGENASE (EC 1.3.1.12) (PDH).	66	50	624
372	3	2150	1599	gi|467416	unknown [Bacillus subtilis]	66	38	552
378	1	212	1009	gi|147309	purine nucleoside phosphorylase [Escherichia coli]	66	50	798
401	1	1	462	gi|388263	p-aminobenzoic acid synthase [Streptomyces griseus] pir|JN0531|JN0531 p-aminobenzoic acid synthase - Streptomyces riseus	66	46	462
404	7	4826	5254	gi|606744	cytidine deaminase [Bacillus subtilis]	66	51	429
411	2	1738	1103	gi|1460081	unknown [Mycobacterium tuberculosis]	66	44	636
420	1	2	541	gi|1046024	Na+ ATPase subunit J [Mycoplasma genitalium]	66	49	540
431	1	1	858	gi|1500008	M. jannaschii predicted coding region MJ1154 [Methanococcus jannaschii]	66	50	858
443	7	5679	5299	gi|852076	HrgA [Bacillus subtilis]	66	46	381
444	3	3405	2413	gi|153047	lysostaphin (ttg start codon) [Staphylococcus simulans] pir|A25881|A25881 lysostaphin precursor - Staphylococcus simulans sp|P10547|LSTP_STASI LYSOSTAPHIN PRECURSOR (EC 3.5.1.-).	66	51	993
561	1	956	480	gi|1204905	DNA-3-methyladenine glycosidase I [Haemophilus influenzae]	66	45	477
562	3	1066	1383	gi|1046082	M. genitalium predicted coding region MG372 [Mycoplasma genitalium]	66	52	318
576	1	11	724	gi|305014	ORF_o234 [Escherichia coli]	66	43	714
577	3	1190	903	gi|1001353	hypothetical protein [Synechocystis sp.]	66	52	288
584	1	2	331	sp|P24204|YEBA_	HYPOTHETICAL 46.7 KD PROTEIN IN HSBB-RUVB INTERGENIC REGION (ORFU).	66	48	330
592	1	1410	706	gi|928839	ORF266; putative [Lactococcus lactis phage BK5-T]	66	51	705
601	1	1433	720	gi|1488695	novel antigen; orf-2 [Staphylococcus aureus]	66	55	714

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
619	3	468	845	gi|746573	similar to M. musculus transport system membrane protein, Nramp PIR:A40739) and S. cerevisiae SMF1 protein (PIR:A45154) Caenorhabditis elegans)	66	45	378
706	2	561	355	gi|804808	unknown protein (Rattus norvegicus)	66	46	207
734	2	673	512	gi|1519085	phosphatidylcholine binding immunoglobulin heavy chain IgM variable region (Mus musculus)	66	60	162
740	1	3	317	gi|1209272	argininosuccinate lyase (Campylobacter jejuni)	66	47	315
764	1	310	747	gi|435296	alkaline phosphatase like protein (Lactococcus lactis) pir|S39339|S39339 alkaline phosphatase-like protein - Lactococcus actis	66	42	438
852	1	338	171	gi|536955	CG Site No. 361 (Escherichia coli)	66	43	168
886	1	3	158	gi|289272	ferrichrome-binding protein (Bacillus subtilis)	66	44	156
889	1	462	232	gi|833061	HCMVUL77 (AA 1-642) (Human cytomegalovirus)	66	66	231
893	1	2	247	gi|149008	putative (Helicobacter pylori)	66	45	246
900	1	1425	733	gi|580842	F3 (Bacillus subtilis)	66	51	693
906	2	2300	1473	gi|790945	aryl-alcohol dehydrogenase (Bacillus subtilis)	66	53	828
947	1	79	549	gi|410117	diaminopimelate decarboxylase (Bacillus subtilis)	66	47	471
950	1	1100	552	gi|48713	orf145 (Staphylococcus aureus)	66	35	549
955	2	89	475	gi|1204390	uridine kinase (uridine monophosphokinase) (Haemophilus influenzae)	66	50	387
981	2	1308	997	gi|457146	rhoptry protein (Plasmodium yoelii)	66	38	312
986	1	25	315	gi|305002	ORF_f356 (Escherichia coli)	66	31	291
1057	1	3	203	gi|1303853	YqgF (Bacillus subtilis)	66	40	201
1087	1	1	294	gi|575913	unknown (Saccharomyces cerevisiae)	66	53	294
1105	1	1	231	gi|1045799	methylgalactoside permease ATP-binding protein (Mycoplasma genitalium)	66	46	231
1128	1	2	574	gi|1001493	hypothetical protein (Synechocystis sp.)	66	46	573
1150	1	498	250	gi|1499034	M. jannaschii predicted coding region MJ0255 (Methanococcus jannaschii)	66	40	249
1180	2	707	453	gi|215908	DNA polymerase (g43) (Bacteriophage T4)	66	46	255
1208	1	1123	587	gi|1256653	DNA-binding protein (Bacillus subtilis)	66	58	537
1342	1	1	402	gi|1208474	hypothetical protein (Synechocystis sp.)	66	53	402
1761	2	589	398	gi|215811	tail fiber protein (Bacteriophage T3)	66	50	192

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim.	% ident	length (nt)
1983	1	499	251	gi|1045935	DNA helicase II (Mycoplasma genitalium)	66	40	249
2103	2	176	400	gi|929798	precursor for the major merozoite surface antigens (Plasmodium alciparum)	66	46	225
2341	1	373	188	gi|1256623	exodeoxyribonuclease (Bacillus subtilis)	66	38	186
2458	1	325	164	gi|1019410	unknown (Schizosaccharomyces pombe)	66	47	162
2505	1	468	235	gi|1510394	putative transcriptional regulator (Methanococcus jannaschii)	66	39	234
2525	1	558	280	gi|1000695	cytotoxin L (Clostridium sordellii)	66	44	279
2935	1	3	275	gi|765073	autolysin (Staphylococcus aureus)	66	47	273
3005	1	114	305	gi|1205784	heterocyst maturation protein (Haemophilus influenzae)	66	46	192
3048	1	80	277	gi|1303813	YqeW (Bacillus subtilis)	66	42	198
3071	1	1	189	gi|1070014	protein-dependent (Bacillus subtilis)	66	41	189
3081	1	404	225	gi|984212	unknown (Schizosaccharomyces pombe)	66	44	180
3090	2	580	386	gi|1204987	DNA polymerase III, alpha chain (Haemophilus influenzae)	66	48	195
3318	1	1	387	gi|1009366	Respiratory nitrate reductase (Bacillus subtilis)	66	49	387
3739	1	798	400	gi|1109684	ProV (Bacillus subtilis)	66	47	399
3796	1	402	202	gi|853760	acyl-CoA dehydrogenase (Bacillus subtilis)	66	60	201
1924	1	525	347	gi|563952	gluconate permease (Bacillus licheniformis)	66	46	249
4240	1	3	350	gi|151259	HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas mevalonii) pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	66	51	348
4604	1	7	234	pir|A26713|BHHC	hemocyanin subunit II - Atlantic horseshoe crab	66	46	228
4	9	8845	9750	gi|145646	cynR (Escherichia coli)	65	35	906
6	5	2708	3565	gi|887824	ORF_o310 (Escherichia coli)	65	47	858
13	1	1993	998	gi|143402	recombination protein (ttg start codon) (Bacillus subtilis) gi|1303923 RecN (Bacillus subtilis)	65	44	996
15	7	2493	3524	gi|1403126	cxcD gene product (Alcaligenes eutrophus)	65	38	1032
18	3	1908	1372	gi|349187	acyltransferase (Saccharomyces cerevisiae)	65	50	537
21	3	1467	2492	gi|149518	phosphoribosyl anthranilate transferase (Lactococcus lactis) pir|S35126|S35126 anthranilate phosphoribosyltransferase (EC .4.2.18) - Lactococcus lactis subsp. lactis	65	52	1026
25	4	3374	4312	gi|1502420	malonyl-CoA:Acyl carrier protein transacylase (Bacillus subtilis)	65	44	939

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt.)
27	2	390	626	gi|1212729	YqhJ (Bacillus subtilis)	65	45	237
31	12	11040	10387	gi|509245	D-hydroxyisocaproate dehydrogenase (Lactobacillus delbrueckii)	65	41	654
38	24	19172	19528	gi|547:	H-protein (Flaveria cronquistii)	65	41	357
44	2	790	1746	gi|405882	yeiK (Escherichia coli)	65	46	957
44	12	9356	8832	gi|1205905	molybdenum cofactor biosynthesis protein (Haemophilus influenzae)	65	50	525
45	8	6635	7588	gi|493074	ApbA protein (Salmonella typhimurium)	65	46	954
51	2	580	1503	gi|580897	OppB gene product (Bacillus subtilis)	65	45	924
52	1	225	953	gi|1205518	NAD(P)H-flavin oxidoreductase (Haemophilus influenzae)	65	45	729
55	4	1339	1058	pir|A44459|A444	troponin T beta TnT-5 - rabbit	65	41	282
67	9	7421	8272	gi|143607	sporulation protein (Bacillus subtilis)	65	42	852
73	5	4446	5375	gi|1204896	lysophospholipase L2 (Haemophilus influenzae)	65	37	930
74	1	954	478	gi|1204844	H. influenzae predicted coding region HI0594 (Haemophilus influenzae)	65	50	477
77	1	2	757	gi|1046082	M. genitalium predicted coding region MG372 (Mycoplasma genitalium)	65	46	756
77	2	795	1433	gi|1222116	permease (Haemophilus influenzae)	65	37	639
81	3	4728	3454	gi|1001708	hypothetical protein (Synechocystis sp.)	65	49	1275
91	7	8548	8357	gi|1399263	cystathionine beta-lyase (Emericella nidulans)	65	40	192
98	3	1608	1988	gi|467423	unknown (Bacillus subtilis)	65	38	381
98	4	2250	2987	gi|467424	unknown (Bacillus subtilis)	65	45	738
102	3	2598	2119	gi|1511532	N-terminal acetyltransferase complex, subunit ARD1 (Methanococcus jannaschii)	65	39	480
102	4	3647	2862	gi|1204637	H. influenzae predicted coding region HI0388 (Haemophilus influenzae)	65	32	786
103	9	10851	9841	gi|142695	S-adenosyl-L-methionine:uroporphyrinogen III methyltransferase Bacillus megaterium)	65	47	1011
103	10	10439	10119	gi|710021	nitrite reductase (nirD) (Bacillus subtilis)	65	51	321
106	2	262	1140	gi|39881	ORF 311 (AA 1-311) (Bacillus subtilis)	65	44	879
109	5	3909	4268	gi|1204399	glucosamine-6-phosphate deaminase protein (Haemophilus influenzae)	65	44	360
109	10	7165	8595	gi|536955	CG Site No. 361 (Escherichia coli)	65	41	1431

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
110	4	3688	3915	gi|407881	stringent response-like protein [Streptococcus equisimilis] pir|S39975|S39975 stringent response-like protein - Streptococcus equisimilis	65	45	228
110	5	3882	4295	gi|407880	ORF1 [Streptococcus equisimilis]	65	50	414
110	6	4231	4380	gi|1139574	Orf2 [Streptomyces griseus]	65	56	150
112	10	9218	8640	gi|1204571	H. influenzae predicted coding region HI0318 [Haemophilus influenzae]	65	52	579
112	12	12049	11288	gi|710496	transcriptional activator protein [Bacillus brevis]	65	32	762
125	1	2	202	gi|1151158	repeat organellar protein [Plasmodium chabaudi]	65	39	201
126	1	3	422	gi|37589	precursor [Homo sapiens]	65	46	420
127	11	10733	12658	gi|1064809	homologous to sp:HTRA_ECOLI [Bacillus subtilis]	65	41	1926
143	8	7543	7004	gi|216513	mutator mutT (AT-GC transversion) [Escherichia coli]	65	56	540
145	5	3587	3838	gi|1209768	D02_orf569 [Mycoplasma pneumoniae]	65	27	252
150	4	3482	2841	gi|1146225	putative [Bacillus subtilis]	65	37	642
166	1	3858	1948	gi|148304	beta-1,4-N-acetylmuramoylhydrolase [Enterococcus hirae] pir|A42296|A42296 lysozyme 2 (EC 3.2.1.-) precursor - Enterococcus irae (ATCC 9790)	65	50	1911
188	6	3195	4178	gi|151943	ORF3; putative [Rhodobacter capsulatus]	65	46	984
189	9	4982	4785	gi|558812	ORF IV (AA 1-489) [Figwort mosaic virus]	65	40	198
195	6	7900	5272	gi|145220	alanyl-tRNA synthetase [Escherichia coli]	65	49	2637
195	7	10599	8104	gi|882711	exonuclease V alpha-subunit [Escherichia coli]	65	38	2496
206	16	16896	18191	gi|408115	ornithine acetyltransferase [Bacillus subtilis]	65	53	1296
217	4	3844	3215	gi|1205974	5'guanylate kinase [Haemophilus influenzae]	65	41	630
220	4	5265	3751	gi|580920	rodD (gtaA) polypeptide (AA 1-673) [Bacillus subtilis] pir|S06048|S06048 probable rodD protein - Bacillus subtilis sp|P13484|TAGF_BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) LPHA-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHOIC ACID BIOSYNTHESIS ROTEIN E).	65	40	1515
236	5	2327	3709	gi|1146200	DNA or RNA helicase, DNA-dependent ATPase [Bacillus subtilis]	65	46	1383
237	3	1902	2513	gi|149379	HisBd [Lactococcus lactis]	65	46	612
241	4	4968	4195	gi|1205308	ribonuclease HII (EC 31264) (RNASE HII) [Haemophilus influenzae]	65	50	774
252	1	1278	940	gi|1204989	hypothetical protein (GB:U00022_9) [Haemophilus influenzae]	65	40	339
261	5	4780	3794	gi|145927	fecD [Escherichia coli]	65	43	987

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match Accession	match gene name	% sim	% ident	length (nt)
274	1	3	278	gi|496558	orfX [Bacillus subtilis]	65	42	276
301	2	982	815	gi|467418	unknown [Bacillus subtilis]	65	45	168
307	4	3586	2864	gi|1070014	protein-dependent [Bacillus subtilis]	65	40	723
335	2	2286	1399	gi|146913	N-acetylglucosamine transport protein [Escherichia coli] pir|B29895|HQEC2N phosphotransferase system enzyme II (EC .7.1.69), N-acetylglucosamine-specific - Escherichia coli sp|P09323|PTAA_ECOLI PTS SYSTEM, N-ACETYLGLUCOSAMINE-SPECIFIC IIABC OMPONENT (EIIA	65	50	888
338	5	4120	3170	gi|1277029	biotin synthase [Bacillus subtilis]	65	49	951
343	3	1490	2800	gi|143264	membrane-associated protein [Bacillus subtilis]	65	48	1311
344	4	2761	2531	gi|1050540	tRNA-glutamine synthetase [Lupinus luteus]	65	34	231
358	3	3421	3621	gi|1146220	NAD+ dependent glycerol-3-phosphate dehydrogenase [Bacillus subtilis]	65	47	201
364	1	238	699	gi|1340128	ORF1 [Staphylococcus aureus]	65	51	462
379	1	1	576	gi|143331	alkaline phosphatase regulatory protein [Bacillus subtilis] pir|A27650|A27650 regulatory protein phoR - Bacillus subtilis sp|P23545|PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN HOR (EC 2.7.3.-).	65	40	576
379	3	3666	4346	gi|143268	dihydrolipoamide transsuccinylase (odhB; EC 2.3.1.61) [Bacillus ubtilis]	65	50	681
428	1	187	483	gi|1420465	ORF YOR195w [Saccharomyces cerevisiae]	65	45	297
438	2	272	838	gi|143498	degS protein [Bacillus subtilis]	65	38	567
444	11	9280	10215	gi|1204756	ribokinase [Haemophilus influenzae]	65	47	936
449	2	1241	1531	gi|599848	Na/H antiporter homolog [Lactococcus lactis]	65	41	291
478	2	1452	865	gi|1045942	glycyl-tRNA synthetase [Mycoplasma genitalium]	65	39	588
479	1	1032	517	gi|1498192	putative [Pseudomonas aeruginosa]	65	40	516
480	6	4312	5637	gi|415662	UDP-N-acetylglucosamine 1-carboxyvinyl transferase [Acinetobacter alcoaceticus]	65	48	1326
484	1	2	430	gi|146551	transmembrane protein (kdpD) [Escherichia coli]	65	44	429
499	1	54	932	gi|603456	reductase [Leishmania major]	65	53	879
505	1	914	459	gi|1518853	OafA [Salmonella typhimurium]	65	39	456
571	2	1509	883	gi|49399	open reading frame upstream glnE [Escherichia coli] ir|S37754|S37754 hypothetical protein XE (glnE 5' region) - cherichia coli	65	44	627
611	2	506	270	gi|10961	RAP-2 [Plasmodium falciparum]	65	40	237

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
705	1	564	283	gi|710020	nitrite reductase (nirB) [Bacillus subtilis]	65	52	282
712	1	1	177	gi|289272	ferrichrome-binding protein [Bacillus subtilis]	65	37	177
712	2	196	354	gi|289272	ferrichrome-binding protein [Bacillus subtilis]	65	37	159
743	1	2	631	gi|310631	ATP binding protein [Streptococcus gordonii]	65	45	630
749	2	393	779	gi|467374	single strand DNA binding protein [Bacillus subtilis]	65	29	387
762	1	1698	850	gi|160399	multidrug resistance protein [Plasmodium falciparum]	65	48	849
788	1	85	315	gi|1129096	unknown protein [Bacillus sp.]	65	35	231
850	1	1	408	gi|1006604	hypothetical protein [Synechocystis sp.]	65	37	408
908	1	1	444	gi|1199546	2362 [Saccharomyces cerevisiae]	65	46	444
925	1	1	174	gi|1256653	DNA-binding protein [Bacillus subtilis]	65	54	174
1031	1	26	232	gi|238657	AppC=cytochrome d oxidase, subunit I homolog [Escherichia coli, K12, eptide, 514 aa]	65	47	207
1037	1	414	262	gi|1491813	gamma-glutamyltranspeptidase [Bacillus subtilis]	65	46	153
1053	1	348	175	gi|642655	unknown [Rhizobium meliloti]	65	34	174
1149	1	1399	752	gi|1162980	ribulose-5-phosphate 3-epimerase [Spinacia oleracea]	65	48	648
1214	1	881	495	gi|1205959	lactam utilization protein [Haemophilus influenzae]	65	45	387
1276	1	476	276	pir|S35493|S354	site-specific DNA-methyltransferase StsI (EC 2.1.1.-) - Streptococcus sanguis	65	35	201
1276	2	900	577	gi|473794	'ORF' [Escherichia coli]	65	34	324
2057	1	272	138	gi|633699	TrsH [Yersinia enterocolitica]	65	21	135
2521	1	336	169	gi|1045789	hypothetical protein (GB:U14003_76) [Mycoplasma genitalium]	65	41	168
2974	1	590	297	gi|152052	enantiomerase-selective amidase [Rhodococcus sp.]	65	45	294
3031	1	306	154	pir|JQ1024|JQ10	hypothetical 30K protein (DmRP140 5' region) - fruit fly (Drosophila melanogaster)	65	45	153
3069	1	3	278	gi|144906	product homologous to E.coli thioredoxin reductase: J.Biol.Chem. 1988) 263:9015-9019, and to F52a protein of alkyl hydroperoxide eductase from S.typhimurium: J.Biol.Chem. (1990) 265:10535-10540; pen reading frame A [Clostridium pasteurianum]	65	46	276
3146	1	282	142	gi|49315	ORF1 gene product [Bacillus subtilis]	65	47	141
3170	1	679	341	gi|1507711	indolepyruvate decarboxylase [Erwinia herbicola]	65	44	339

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
3546	1	1	303	gi|450688	hsdM gene of EcoprrI gene product [Escherichia coli] pir|S38437|S38437 hsdM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	65	42	303
3782	1	2	328	gi|166412	NADH-glutamate synthase [Medicago sativa]	65	42	327
3990	1	374	189	gi|1009366	Respiratory nitrate reductase [Bacillus subtilis]	65	53	186
4032	1	613	308	gi|1323127	ORF YGR087c [Saccharomyces cerevisiae]	65	50	306
4278	2	726	364	gi|1197667	vitellogenin [Anolis pulchellus]	65	42	363
19	4	4259	5518	gi|145727	deaD [Escherichia coli]	64	45	1260
19	6	7639	6926	gi|1016232	ycf27 gene product [Cyanophora paradoxa]	64	36	714
20	8	7053	6454	gi|765073	autolysin [Staphylococcus aureus]	64	47	600
31	13	12706	11537	gi|414009	ipa-85d gene product [Bacillus subtilis]	64	45	1170
33	4	2388	4364	gi|1204696	fructose-permease IIBC component [Haemophilus influenzae]	64	47	1977
36	3	1871	3013	gi|290503	glutamate permease [Escherichia coli]	64	40	1143
37	6	4065	4409	gi|39815	orf 2 gene product [Bacillus subtilis]	64	46	345
45	9	7852	8760	gi|1230585	nucleotide sugar epimerase [Vibrio cholerae 0139]	64	53	909
53	3	1540	1899	gi|1303961	YqjJ [Bacillus subtilis]	64	50	360
56	6	4793	3855	gi|457514	gltC [Bacillus subtilis]	64	45	939
56	24	30002	30247	gi|470331	similar to zinc fingers [Caenorhabditis elegans]	64	42	246
62	4	2759	2421	gi|642655	unknown [Rhizobium meliloti]	64	28	339
85	6	7178	6027	gi|457702	5-aminoimidazole ribonucleotide-carboxilase [Pichia methanolica] pir|S39112|S39112 phosphoribosylaminoimidazole carboxylase (EC .1.1.21) - yeast [Pichia methanolica]	64	46	1152
96	9	9251	10030	gi|1511513	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	64	42	780
100	1	1	600	gi|765073	autolysin [Staphylococcus aureus]	64	44	600
106	5	3868	4854	gi|466778	lysine specific permease [Escherichia coli]	64	46	987
123	2	838	554	gi|467484	unknown [Bacillus subtilis]	64	47	285
127	8	7514	7810	gi|210061	serotype-specific antigen [African horse sickness virus] pir|S27891|S27891 capsid protein VP2 - African horse sickness virus	64	28	297
131	7	7134	6721	gi|1511160	M. jannaschii predicted coding region MJ1163 [Methanococcus jannaschii]	64	46	414

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
142	5	5455	4817	gi|1173517	riboflavin synthase alpha subunit [Actinobacillus pleuropneumoniae]	64	44	639
143	1	709	356	pir|A32950|A329	probable reductase protein - Leishmania major	64	52	354
149	10	3555	3295	gi|398151	major surface antigen MSG2 [Pneumocystis carinii]	64	46	261
154	4	3134	2307	gi|984587	DinP [Escherichia coli]	64	50	828
161	5	3855	4880	gi|903304	ORF72 [Bacillus subtilis]	64	37	1026
165	1	33	791	gi|467483	unknown [Bacillus subtilis]	64	38	759
175	6	6355	4714	gi|1072398	phaD gene product [Rhizobium meliloti]	64	42	1512
188	3	2042	2500	gi|1001961	MHC class II analog [Staphylococcus aureus]	64	45	459
195	14	13667	13446	gi|396380	No definition line found [Escherichia coli]	64	47	222
206	15	16429	16938	gi|304134	argC [Bacillus stearothermophilus]	64	49	510
215	1	560	282	gi|142359	ORF 6 [Azotobacter vinelandii]	64	39	279
243	7	7818	6928	gi|414014	ipa-90d gene product [Bacillus subtilis]	64	49	891
258	2	1330	845	gi|664754	P17 [Listeria monocytogenes]	64	38	486
259	1	462	232	gi|1499663	M. jannaschii predicted coding region MJ0837 [Methanococcus jannaschii]	64	52	231
263	6	6565	5567	gi|142828	aspartate semialdehyde dehydrogenase [Bacillus subtilis] sp|Q04797|DHAS_BACSU ASPARTATE-SEMIALDEHYDE DEHYDROGENASE (EC .2.1.11) (ASA DEHYDROGENASE).	64	48	999
271	1	3	1163	gi|467091	hflX; B2235_C2_202 [Mycobacterium leprae]	64	44	1161
280	1	173	1450	gi|1303839	YqfR [Bacillus subtilis]	64	43	1278
293	1	2532	1267	gi|147345	primosomal protein n' [Escherichia coli]	64	45	1266
295	2	742	1488	gi|459266	Potential membrane spanning protein [Staphylococcus hominis] pir|S42932|S42932 potential membrane spanning protein - taphylococcus hominis	64	39	747
301	5	1625	1446	gi|580835	lysine decarboxylase [Bacillus subtilis]	64	35	180
315	4	5064	3949	gi|143396	quinol oxidase [Bacillus subtilis]	64	45	1116
321	1	1264	635	gi|710496	transcriptional activator protein [Bacillus brevis]	64	41	630
333	5	4520	4239	gi|1314295	ORF2; putative 19 kDa protein [Listeria monocytogenes]	64	43	282
342	1	1	549	gi|142940	ftsA [Bacillus subtilis]	64	38	549
353	3	2878	2324	gi|537049	ORF_o470 [Escherichia coli]	64	44	555

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
379	2	827	3658	|pir|S25295|A328	oxoglutarate dehydrogenase (lipoamide) (EC 1.2.4.2) - Bacillus subtilis	64	47	2832
404	6	4429	4839	|pir|A36933|A369	diacylglycerol kinase homolog - Streptococcus mutans	64	35	411
407	1	2020	1133	|gi|969026	OrfX (Bacillus subtilis)	64	41	888
425	1	1109	591	|gi|1146177	phosphotransferase system glucose-specific enzyme II (Bacillus subtilis)	64	44	519
443	6	4082	4798	|gi|147309	purine nucleosida phosphorylase (Escherichia coli)	64	51	717
450	2	1035	1604	|gi|606376	ORF_o162 (Escherichia coli)	64	38	570
470	5	1680	6107	|gi|1369948	host interacting protein (Bacteriophage B1)	64	45	4428
486	4	1911	1471	|gi|1205582	spermidine/putrescine transport system permease protein (Haemophilus influenzae)	64	35	441
497	1	2217	1159	|sp|P36929|FMU_E	FMU PROTEIN.	64	38	1059
501	1	3	410	|gi|142450	ahrC protein (Bacillus subtilis)	64	38	408
514	1	3	290	|gi|1204496	H. influenzae predicted coding region HI0238 (Haemophilus influenzae)	64	34	288
551	4	3162	3323	|gi|1204511	bacterioferritin comigratory protein (Haemophilus influenzae)	64	41	162
603	4	759	956	|gi|755823	NADH dehydrogenase F (Streptogyna americana)	64	35	198
653	2	940	746	|gi|.213234	dicarboxylic amino acids DipSp permease (Saccharomyces cerevisiae)	64	41	195
660	3	3H01	2257	|sp|P46133|YDAH_	HYPOTHETICAL PROTEIN IN OGT 5'REGION (FRAGMENT).	64	39	1545
695	1	11	502	|gi|1001383	hypothetical protein (Synechocystis sp.)	64	41	492
702	1	3	752	|gi|142865	DNA primase (Bacillus subtilis)	64	46	750
826	1	1	339	|gi|971336	arginyl tRNA synthetase (Bacillus subtilis)	64	50	339
838	1	1831	917	|gi|1354775	pfoS/R (Treponema pallidum)	64	41	915
864	3	675	944	|gi|39833	cyclomaltodextrin glucanotransferase (Bacillus stearothermophilus) i|39835 cyclomaltodextrin glucanotransferase (Bacillus earothermophilus)	64	47	270
887	1	3	677	|gi|153002	enterotoxin type E precursor (Staphylococcus aureus) pir|A28179|A28179 enterotoxin E precursor - Staphylococcus aureus sp|P12993|ETXE_STAAU ENTEROTOXIN TYPE E PRECURSOR (SEE).	64	46	675
928	2	1172	963	|gi|311976	fibrinogen-binding protein (Staphylococcus aureus) pir|S34270|S34270 fibrinogen-binding protein - Staphylococcus ureus	64	41	210
1049	2	800	606	|gi|1049115	Rap60 (Bacillus subtilis)	64	42	195
1067	2	999	748	|gi|1151072	HhdA precursor (Haemophilus ducreyi)	64	50	252

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
1120	1	50	202	gi|142439	ATP-dependent nuclease [Bacillus subtilis]	64	30	153
1125	1	751	377	gi|581648	epiB gene product [Staphylococcus epidermidis]	64	44	375
1688	1	402	214	pir|A01365|TVMS	transforming protein K-ras - mouse	64	47	189
2472	1	2	358	gi|487282	Na+ -ATPase subunit J [Enterococcus hirae]	64	36	357
2989	1	520	356	gi|304134	argC [Bacillus stearothermophilus]	64	50	165
3013	1	630	352	gi|551699	cytochrome oxidase subunit I [Bacillus firmus]	64	51	279
3034	1	546	274	gi|1204349	hypothetical protein (GB:GB:D90212_3) [Haemophilus influenzae]	64	50	273
3197	1	613	308	gi|1009366	Respiratory nitrate reductase [Bacillus subtilis]	64	46	306
3303	1	90	362	gi|1107839	alginate lyase [Pseudomonas aeruginosa]	64	43	273
3852	2	82	288	gi|216746	D-lactate dehydrogenase [Lactobacillus plantarum]	64	42	207
3868	1	1	312	gi|149435	putative [Lactococcus lactis]	64	48	312
3918	1	660	331	gi|5532	acetyl-CoA acyltransferase [Yarrowia lipolytica]	64	46	330
4000	1	112	378	gi|984688	unknown [Saccharomyces cerevisiae]	64	44	367
4009	1	81	368	gi|39372	grsB gene product [Bacillus brevis]	64	41	288
4166	1	2	349	gi|149435	putative [Lactococcus lactis]	64	46	348
4366	1	2	307	gi|216267	ORF2 [Bacillus megaterium]	64	44	306
4457	1	2	400	gi|1197667	vitellogenin [Anolis pulchellus]	64	43	399
11	3	1539	2438	gi|438228	ORF C [Staphylococcus aureus]	63	32	900
24	7	5611	5423	gi|1369943	a1 gene product [Bacteriophage B1]	63	34	189
29	1	1	390	gi|467441	expressed at the end of exponential growyh under condtions in which he enzymes of the TCA cycle are repressed [Bacillus subtilis] gi|467441 expressed at the end of exponential growyh under ondtions in which the enzymes of the TCA cycle are repressed Bacil	63	43	390
31	6	6329	5712	gi|496943	ORF [Saccharomyces cerevisiae]	63	47	618
44	23	14669	15019	pir|A04446|QQEC	hypothetical protein F-92 - Escherichia coli	63	36	351
48	6	4403	6250	gi|43498	pyruvate synthase [Halobacterium halobium]	63	42	1848
50	5	3869	4738	gi|413967	ipa-43d gene product [Bacillus subtilis]	63	43	870
53	6	6764	5742	gi|474176	regulator protein [Staphylococcus xylosus]	63	49	1023

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
56	14	15880	17607	gi|467409	DNA polymerase III subunit (Bacillus subtilis)	63	44	1728
57	11	7945	7376	gi|537036	ORF_o158 (Escherichia coli)	63	39	570
62	3	2479	2114	gi|142656	unknown (Rhizobium meliloti)	63	41	366
70	8	6562	7353	gi|1399821	PhoC (Rhizobium meliloti)	63	46	792
75	2	223	927	gi|149376	HisG (Lactococcus lactis)	63	45	705
78	5	4912	4403	gi|413950	ipa-26d gene product (Bacillus subtilis)	63	42	510
91	5	9076	7220	gi|466997	metH2; B2126_C1_157 (Mycobacterium leprae)	63	41	1857
91	8	10566	9448	gi|1204344	cystathionine gamma-synthase (Haemophilus influenzae)	63	45	1119
120	1	21	1508	gi|882657	sulfite reductase (NADPH) flavoprotein beta subunit (Escherichia oli)	63	46	1488
120	4	2722	4125	gi|665994	hypothetical protein (Bacillus subtilis)	63	34	1404
127	7	6064	7566	gi|40162	murE gene product (Bacillus subtilis)	63	44	1503
149	6	2321	2106	gi|148503	dnaK (Erysipelothrix rhusiopathiae)	63	40	216
149	26	10445	10170	gi|4870	ORF 2, has similarity to DNA polymerase (Saccharomyces kluyveri) r|S15961|S15961 hypothetical protein 2 - yeast (Saccharomyces yveri) plasmid pSKL	63	42	276
164	2	507	1298	gi|145476	CDP-diglyceride synthetase (Escherichia coli)	63	44	792
166	6	9909	8164	gi|151932	fructose enzyme II (Rhodobacter capsulatus)	63	41	1746
169	4	1704	1886	gi|152886	elongation factor Ts (tsf) (Spiroplasma citri)	63	48	183
188	5	3145	2951	gi|1334547	GIY COI i14 grp IB protein (Podospora anserina)	63	42	195
195	13	11767	12804	gi|606100	ORF_o335 (Escherichia coli)	63	40	1038
201	2	607	2283	gi|433534	arginyl-tRNA synthetase (Corynebacterium glutamicum) pir|A49936|A49936 arginine--tRNA ligase (EC 6.1.1.19) - orynebacterium glutamicum	63	46	1677
206	14	15893	16489	gi|580828	N-acetyl-glutamate-gamma-semialdehyde dehydrogenase (Bacillus ubtilis)	63	49	597
220	5	7769	5766	gi|216334	secA protein (Bacillus subtilis)	63	42	2004
221	1	74	907	gi|677945	AppA (Bacillus subtilis)	63	42	834
227	3	944	1708	gi|1510558	cobyric acid synthase (Methanococcus jannaschii)	63	46	765
261	2	804	1070	gi|486511	ORF YKR054c (Saccharomyces cerevisiae)	63	45	267
269	2	3606	1960	gi|148221	DNA-dependent ATPase, DNA helicase (Escherichia coli) pir|JS0137|BVECRQ recQ protein - Escherichia coli	63	42	1647

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
278	8	7417	6176	|gi|o99273	|cystathionine gamma-synthase (Mycobacterium leprae) sp|P46807|METB_MYCLE CYSTATHIONINE GAMMA-SYNTHASE (EC 4.2.99.9) O-SUCCINYLHOMOSERINE (THIOL)-LYASE).	63	41	1242
287	2	738	1733	|gi|405133	|putative (Bacillus subtilis)	63	38	996
295	1	2	748	|gi|1239983	|hypothetical protein (Bacillus subtilis)	63	41	747
328	3	2148	3134	|gi|45302	|carrier protein (AA 1 - 437) (Pseudomonas aeruginosa) ir|S11497|S11497 branched-chain amino acid tr. .sport protein braB - eudomonas aeruginosa	63	36	987
362	2	1026	1216	|sp|P35136|SERA_	|D-3-PHOSPHOGLYCERATE DEHYDROGENASE (EC 1.1.1.95) (PGDH).	63	38	411
404	1	326	1051	|gi|1303816	|YqeZ (Bacillus subtilis)	63	35	726
405	3	2101	1715	|gi|1303914	|YqhY (Bacillus subtilis)	63	42	387
406	1	451	227	|gi|142152	|sulfate permease (gtg start codon) (Synechococcus PCC6301) pir|A30301|GRYCS7 sulfate transport protein - Synechococcus sp. PCC 7942)	63	43	225
415	2	1048	2718	|gi|1205402	|transport ATP-binding protein (Haemophilus influenzae)	63	41	1671
426	4	3575	2679	|gi|393268	|29-kiloDalton protein (Streptococcus pneumoniae) sp|P42362|P29K_STRPN 29 KD MEMBRANE PROTEIN IN PSAA 5'REGION ORF1).	63	39	897
505	3	1347	2195	|gi|1418999	|orf4 (Lactobacillus sake)	63	40	849
507	1	2	574	|gi|546917	|comK (Bacillus subtilis, E26, Peptide, 192 aa)	63	35	573
562	2	146	1084	|gi|43985	|nifS-like gene (Lactobacillus delbrueckii)	63	45	939
675	1	427	215	|gi|1510994	|serine aminotransferase (Methanococcus jannaschii)	63	29	213
686	1	3	230	|gi|517356	|nitrate reductase (NADH) (Lotus japonicus)	63	52	228
701	1	3	392	|gi|881940	|NorQ protein (Paracoccus denitrificans)	63	41	390
720	1	2	400	|gi|47168	|open reading frame (Streptomyces lividans)	63	35	399
779	1	571	287	|gi|1261932	|unknown (Mycobacterium tuberculosis)	63	41	285
907	1	22	321	|gi|149445	|ORF1 (Lactococcus lactis)	63	27	300
972	1	794	399	|gi|1511235	|M. jannaschii predicted coding region MJ1232 (Methanococcus jannaschii)	63	27	396
1085	1	1154	618	|gi|1204277	|hypothetical protein (GB:U00019_14) (Haemophilus influenzae)	63	38	537
1094	1	3	542	|gi|790943	|urea amidolyase (Bacillus subtilis)	63	39	540
1108	1	3	482	|pir|S49892|S498	|regulation protein - Bacillus subtilis	63	44	480
1113	1	1231	617	|gi|493017	|endocarditis specific antigen (Enterococcus faecalis)	63	45	615

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
1300	1	3	695	|sp|P33940|YOJH_	|HYPOTHETICAL 54.3 KD PROTEIN IN ECO-ALKB INTERGENIC REGION.	63	46	693
1325	1	1	204	|gi|928989	|p100 protein [Borrelia burgdorferi]	63	30	204
1814	1	3	245	|gi|1303914	|YqhY [Bacillus subtilis]	63	34	243
2021	1	498	250	|pir|C33496|C334	|hisC homolog - Bacillus subtilis	63	46	249
2325	1	2	193	|gi|436132	|product is similar to TnpA of transposon Tn554 from Staphylococcus aureus [Clostridium butyricum]	63	40	192
2335	1	1	195	|gi|1184298	|flagellar MS-ring protein [Borrelia burgdorferi]	63	47	195
2406	1	451	227	|gi|1041785	|rhoptry protein [Plasmodium yoelii]	63	33	225
2961	2	136	360	|gi|312443	|carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus aldolyticus]	63	52	225
2965	1	1	402	|gi|1407784	|orf-1; novel antigen [Staphylococcus aureus]	63	50	402
2987	1	583	293	|gi|1224069	|amidase [Moraxella catarrhalis]	63	35	291
2994	1	266	135	|gi|836646	|phosphoribosylformimino-praic ketoisomerase [Rhodobacter phaeroides]	63	51	132
3043	1	440	252	|gi|1480237	|phenylacetaldehyde dehydrogenase [Escherichia coli]	63	40	189
3078	1	609	400	|gi|1487982	|intrinsic membrane protein [Mycoplasma hominis]	63	36	210
3139	1	2	217	|gi|439126	|glutamate synthase (NADPH) [Azospirillum brasilense] pir|A49916|A49916 glutamate synthase (NADPH) (EC 1.4.1.13) - azospirillum brasilense	63	47	216
1625	1	793	198	|gi|623073	|ORF160; putative [Bacteriophage LL-H]	63	48	896
3658	1	1	399	|gi|1303697	|YrkA [Bacillus subtilis]	63	37	399
3659	1	3	395	|gi|1256135	|YbbF [Bacillus subtilis]	63	48	393
3783	1	720	361	|gi|1256902	|Pyruvate decarboxylase isozyme 2 (Swiss Prot. accession number P16467) [Saccharomyces cerevisiae]	63	34	360
3900	1	338	171	|sp|P10537|AMYB_	|BETA-AMYLASE (EC 3.2.1.2) (1,4-ALPHA-D-GLUCAN MALTOHYDROLASE).	63	54	168
4309	1	3	176	|pir|A37967|A379	|neural cell adhesion molecule Ng-CAM precursor - chicken	63	57	174
4367	1	1	195	|gi|1321932	|Per6p gene product [Pichia pastoris]	63	30	195
4432	1	1	312	|gi|151259	|HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas mevalonii) pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	63	51	312
4468	1	6	308	|gi|296464	|ATPase [Lactococcus lactis]	63	36	303
33	3	1411	2400	|gi|153675	|tagatose 6-P kinase [Streptococcus mutans]	62	44	990
36	9	5985	6218	|gi|1490521	|HHSH3 [Homo sapiens]	62	51	234

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
37	1	2	721	gi|1107531	ceuE gene product (Campylobacter coli)	62	33	720
38	15	10912	11589	gi|1222058	H. influenzae predicted coding region HIN1279 (Haemophilus influenzae)	62	38	678
38	25	19526	20329	gi|695280	ORF2 (Alcaligenes eutrophus)	62	41	804
57	2	2523	1780	gi|171234	orf1 (Haemophilus influenzae)	62	55	744
57	9	6646	6350	gi|508174	EIIB domain of PTS-dependent Gat transport and phosphorylation Escherichia coli	62	35	297
58	1	2	559	gi|755152	highly hydrophobic integral membrane protein (Bacillus subtilis) sp|P42953|TAGG_BACSU TEICHOIC ACID TRANSLOCATION PERMEASE PROTEIN AGG.	62	34	558
67	10	8250	9014	gi|470683	Shows similarity with ATP-binding proteins from other ADC-transport perons, Swiss Prot Accession Numbers P24137, P08007, P04285, P24136 Escherichia coli	62	34	765
69	8	8315	7494	gi|46816	actVA 4 gene product (Streptomyces coelicolor)	62	44	822
80	3	1793	1320	gi|39993	UDP-N-acetylmuramoylalanine--D-glutamate ligase (Bacillus subtilis)	62	43	474
87	7	7034	9205	gi|217191	5'-nucleotidase precursor (Vibrio parahaemolyticus)	62	48	2172
100	3	4051	3089	gi|1511047	phosphoglycerate dehydrogenase (Methanococcus jannaschii)	62	42	963
102	1	2	520	gi|153655	mismatch repair protein (Streptococcus pneumoniae) pir|C28667|C28667 DNA mismatch repair protein hexA - Streptococcus neumoniae	62	34	519
112	2	466	1068	gi|153741	ATP-binding protein (Streptococcus mutans)	62	37	603
114	7	6855	7562	gi|1204866	L-fucose operon activator (Haemophilus influenzae)	62	38	708
116	4	6823	5633	gi|677947	AppC (Bacillus subtilis)	62	37	1191
124	8	6855	6004	gi|853777	product similar to E.coli PRFA2 protein (Bacillus subtilis) pir|S55438|S55438 ywkE protein - Bacillus subtilis sp|P45873|HEMK_BACSU POSSIBLE PROTOPORPHYRINOGEN OXIDASE (EC .3.3.-).	62	44	852
148	1	24	554	gi|467456	unknown (Bacillus subtilis)	62	30	531
149	20	7591	6725	gi|1205807	replicative DNA helicase (Haemophilus influenzae)	62	41	867
161	1	1503	1153	gi|40067	X gene product (Bacillus sphaericus)	62	42	351
164	15	14673	15632	gi|42219	P35 gene product (AA 1 - 314) (Escherichia coli)	62	38	960
165	2	1166	1447	gi|403936	phenylalanyl-tRNA synthetase alpha subunit (Gly294 variant) unidentified cloning vector)	62	38	282
166	2	2084	5089	gi|308861	GTG start codon (Lactococcus lactis)	62	44	3006
171	1	1225	614	gi|1046053	hypothetical protein (SP:P32049) (Mycoplasma genitalium)	62	41	612

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
183	1	2521	1310	gi|143045	hemY (Bacillus subtilis)	62	45	1212
200	1	3	956	gi|142439	ATP-dependent nuclease (Bacillus subtilis)	62	32	954
237	2	935	1966	gi|41695	hisC protein (Escherichia coli)	62	44	1032
261	3	4008	2605	gi|143121	ORF A; putative (Bacillus firmus)	62	42	1404
299	8	4477	4719	gi|467441	expressed at the end of exponential growyh undur condtions in which he enzymes of the TCA cycle are repressed (Bacillus subtilis) gi|467441 expressed at the end of exponential growyh under ondtions in which the enzymes of the TCA cycle are repressed Bacil	62	47	243
304	6	5018	3819	gi|153015	FemA protein (Staphylococcus aureus)	62	43	1200
324	1	2	262	gi|142717	cytochrome aa3 controlling protein (Bacillus subtilis) pir|A33960|A33960 cta protein - Bacillus subtilis sp|P12946|CTAA_BACSU CYTOCHROME AA3 CONTROLLING PROTEIN.	62	30	261
325	2	269	1207	gi|581088	methionyl-tRNA formyltransferase (Escherichia coli)	62	39	939
332	6	4894	4631	gi|1499960	uridine 5'-monophosphate synthase (Methanococcus jannaschii)	62	36	264
355	1	2	370	gi|145925	fecB (Escherichia coli)	62	32	369
365	8	6628	6804	gi|413943	ipa-19d gene product (Bacillus subtilis)	62	54	177
369	2	2744	1626	pir|A43577|A435	regulatory protein pfoR - Clostridium perfringens	62	42	1119
370	1	34	264	gi|40665	beta-glucosidase (Clostridium thermocellum)	62	37	231
415	3	2709	3176	gi|1205401	transport ATP-binding protein (Haemophilus influenzae)	62	35	468
429	1	1578	790	gi|1046024	Na+ ATPase subunit J (Mycoplasma genitalium)	62	40	789
444	2	704	1369	gi|581510	nodulation gene; integral membrane protein; homology to Rhizobium eguminosarum nodI (Rhizobium loti)	62	37	666
477	2	751	1869	pir|A48440|A484	ring-infected erythrocyte surface antigen 2, RESA-2 - Plasmodium falciparum	62	44	1119
485	1	241	1707	gi|17934	betaine aldehyd dehydrogenase (Beta vulgaris)	62	43	1467
487	3	1141	1311	gi|149445	ORF1 (Lactococcus lactis)	62	31	171
494	2	1134	1313	gi|166835	ribulose bisphosphate carboxylase/oxygenase activase (Arabidopsis haliana)	62	37	180
518	1	193	882	gi|153491	O-methyltransferase (Streptomyces glaucescens)	62	39	690
534	2	369	2522	gi|1480429	putative transcriptional regulator (Bacillus stearothermophilus)	62	35	2154
551	6	4371	4820	gi|511113	ferric uptake regulation protein (Campylobacter jejuni)	62	37	450
574	1	1	570	gi|153000	enterotoxin B (Staphylococcus aureus)	62	43	570

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
590	2	344	1171	gi|40367	ORFC [Clostridium acetobutylicum]	62	37	828
655	1	396	830	gi|147195	phnB protein [Escherichia coli]	62	44	435
656	1	2	478	gi|1205451	cell division inhibitor [Haemophilus influenzae]	62	36	477
676	1	692	348	gi|1511613	methyl coenzyme M reductase system, component A2 [Methanococcus jannaschii]	62	36	345
687	1	493	248	gi|49272	Asparaginase [Bacillus licheniformis]	62	48	246
700	2	267	944	gi|1205822	hypothetical protein (GB:X75627_4) [Haemophilus influenzae]	62	40	678
840	2	1715	1041	gi|1045865	M. genitalium predicted coding region MG181 [Mycoplasma genitalium]	62	36	675
864	4	898	1491	gi|1144332	deoxyuridine nucleotidohydrolase [Homo sapiens]	62	38	594
916	1	35	400	gi|413931	ipa-7d gene product [Bacillus subtilis]	62	45	366
1071	1	1	771	gi|1510649	aspartokinase I [Methanococcus jannaschii]	62	40	771
1084	1	19	609	gi|688011	AgX-1 antigen [human, infertile patient, testis, Peptide, 505 aa]	62	39	591
1103	1	3	203	gi|581261	ORF homologous to E.coli motB [Herpetosiphon aurantiacus] pir|S14030|S14030 Hypothetical protein - Herpetosiphon aurantiacus (fragment)	62	51	201
1217	1	463	233	gi|460025	ORF2, putative [Streptococcus pneumoniae]	62	41	231
1533	1	644	414	gi|413968	ipa-44d gene product [Bacillus subtilis]	62	48	231
1537	1	3	257	gi|1510641	alanyl-tRNA synthetase [Methanococcus jannaschii]	62	29	255
2287	1	3	161	gi|485956	mrpC gene product [Proteus mirabilis]	62	45	159
2386	1	3	245	gi|285708	nontoxic component [Clostridium botulinum]	62	31	243
2484	1	331	167	gi|142092	DNA-repair protein (recA) [Anabaena variabilis]	62	35	165
2490	1	798	400	gi|581648	epiB gene product [Staphylococcus epidermidis]	62	42	399
3016	1	596	300	gi|710022	uroporphyrinogen III [Bacillus subtilis]	62	51	297
3116	1	1	213	gi|466883	nifS; B1496_C2_193 [Mycobacterium leprae]	62	44	213
3297	1	823	413	gi|475715	acetyl coenzyme A acetyltransferase (thiolase) [Clostridium cetobutylicum]	62	42	411
3609	1	31	276	gi|1408501 |	homologous to N-acyl-L-amino acid amidohydrolase of Bacillus stearothermophilus [Bacillus subtilis]	62	48	246
3665	2	584	402	gi|151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mevalonii] pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	62	40	183
3733	1	3	374	gi|1353197	thioredoxin reductase [Eubacterium acidaminophilum]	62	42	372

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acesaion	match gene name	% sim	% ident	length (nt)
3898	1	1	237	gi|153675	tagatose 6-P kinase [Streptococcus mutans]	62	45	237
4027	1	283	143	gi|330705	homologue to gene 30 (aa 1-59); putative [Bovine herpesvirus 4]	62	43	141
4109	1	727	365	gi|41748	hsdM protein (AA 1-520) [Escherichia coli]	62	45	363
4303	1	1	303	gi|1303813	YqeW [Bacillus subtilis]	62	43	303
4380	1	530	267	gi|1235684	mevalonate pyrophosphate decarboxylase [Saccharomyces cerevisiae]	62	55	264
4494	1	2	256	gi|510692	enterotoxin H [Staphylococcus aureus]	62	34	255
4598	1	411	223	gi|763513	ORF4; putative [Streptomyces violaceoruber]	62	45	189
4624	1	1	222	gi|41748	hsdM protein (AA 1-520) [Escherichia coli]	62	45	222
5	5	4288	3912	gi|928831	ORF95; putative [Lactococcus lactis phage BK5-T]	61	36	357
11	1	320	162	pir|C33356|C333	prothymosin alpha homolog (clone 32) - human (fragment)	61	33	159
16	11	10991	11938	gi|1205391	hypothetical protein (SP:P33995) [Haemophilus influenzae]	61	44	948
32	1	281	801	gi|1066504	exo-beta 1,3 glucanase [Cochliobolus carbonum]	61	50	519
38	3	616	1107	gi|1510864	glutamine transport ATP-binding protein Q [Methanococcus jannaschii]	61	41	492
45	4	3082	4038	gi|1109686	ProX [Bacillus subtilis]	61	45	957
48	8	7118	7504	gi|498839	ORF2 [Clostridium perfringens]	61	33	387
51	9	4605	5570	gi|388269	traC [Plasmid pAD1]	61	42	966
60	6	1689	2243	gi|1205893	hypothetical protein (GB:U00011_3) [Haemophilus influenzae]	61	32	555
62	9	5559	5122	gi|854656	Na/H antiporter system ORF2 [Bacillus alcalophilus]	61	38	438
67	5	4330	5646	gi|466612	nikA [Escherichia coli]	61	36	1317
74	2	2400	1504	gi|1204846	carbamate kinase [Haemophilus influenzae]	61	40	897
85	1	2198	1101	gi|1498756	amidophosphoribosyltransferase PurF [Rhizobium etli]	61	41	1098
86	4	1995	1582	gi|1499931	M. jannaschii predicted coding region MJ1083 [Methanococcus jannaschii]	61	44	414
97	1	74	649	gi|1518679	orf [Bacillus subtilis]	61	44	576
99	2	2454	1990	gi|413958	ipa-34d gene product [Bacillus subtilis]	61	18	465
124	7	6223	5123	gi|556881	Similar to Saccharomyces cerevisiae SUA5 protein [Bacillus subtilis] pir|S49358|S49358 ipc-29d protein - Bacillus subtilis sp|P39153|YWLC_BACSU HYPOTHETICAL 37.0 KD PROTEIN IN SPOIIR-GLYC NTERGENIC REGION.	61	46	1101
125	4	1668	2531	gi|1491643	ORFA gene product [Chloroflexus aurantiacus]	61	43	864

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
132	1	1250	627	pir|PQ0259|PQ02	hypothetical protein I - Enterococcus faecalis plasmid pAM-beta-1 (fragment)	61	43	624
149	9	3617	3075	gi|1144332	deoxyuridine nucleotidohydrolase (Homo sapiens)	61	40	543
149	22	8690	7869	gi|160047	p101/acidic basic repeat antigen (Plasmodium falciparum) pir|A29232|A29232 101K malaria antigen precursor - Plasmodium alciparum (strain Camp)	61	35	822
168	3	1915	2361	gi|1499694	HIT protein, member of the HIT-family (Methanococcus jannaschii)	61	41	447
171	9	9675	7948	gi|467446	similar to SpoVB (Bacillus subtilis)	61	38	1728
174	3	1042	2340	gi|216374	glutaryl 7-ACA acylase precursor (Bacillus laterosporus)	61	49	1299
190	4	5034	4111	gi|409286	bmrU (Bacillus subtilis)	61	37	924
216	1	2	190	gi|415861	eukaryotic initiation factor 2 beta (eIF-2 beta) (Oryctolagus uniculus)	61	29	189
227	7	4161	5048	gi|216341	ORF for methionine amino peptidase (Bacillus subtilis)	61	41	888
238	4	1959	3047	gi|409543	CbrC protein (Erwinia chrysanthemi)	61	38	1089
247	1	2	694	gi|537231	ORF_f579 (Escherichia coli)	61	38	693
247	2	678	1034	gi|142226	chvD protein (Agrobacterium tumefaciens)	61	40	357
257	2	3523	2627	gi|699379	glvr-1 protein (Mycobacterium leprae)	61	40	897
268	2	3419	3051	gi|40364	ORFA1 (Clostridium acetobutylicum)	61	41	369
275	4	4621	4827	gi|1204848	hypothetical protein (GP:M87049_57) (Haemophilus influenzae)	61	36	207
277	1	1	1845	gi|784897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir|A56390|A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	45	1845
278	9	8003	7032	gi|467462	cysteine synthetase A (Bacillus subtilis)	61	43	972
278	10	9878	8535	gi|1205919	Na+ and Cl- dependent gamma-aminobutryic acid transporter (Haemophilus influenzae)	61	38	1344
283	1	1	366	gi|755607	polyA polymerase (Bacillus subtilis)	61	36	366
288	2	1918	1496	gi|388108	cell wall enzyme (Enterococcus faecalis)	61	43	423
291	1	86	334	gi|454265	FBP3 (Petunia hybrida)	61	38	249
318	1	1104	694	gi|290531	similar to beta-glucoside transport protein (Escherichia coli) sp|P31451|PTIB_ECOLI PTS SYSTEM, ARBUTIN-LIKE IIB COMPONENT PHOSPHOTRANSFERASE ENZYME II, B COMPONENT) (EC 2.7.1.69).	61	47	411
330	2	1912	1190	gi|1001805	hypothetical protein (Synechocystis sp.)	61	41	723

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
385	2	1513	1025	gi|533098	DnaD protein [Bacillus subtilis]	61	42	489
426	1	794	399	gi|1303853	YqgF [Bacillus subtilis]	61	44	396
438	3	810	1421	gi|1293660	AbsA2 [Streptomyces coelicolor]	61	36	612
454	1	1580	792	gi|733522	phosphatidylinositol-4,5-diphosphate 3-kinase [Dictyostelium iscoideum]	61	30	789
464	2	784	560	gi|1123120	C53B7.5 gene product [Caenorhabditis elegans]	61	38	225
470	8	6077	7357	gi|623073	ORF360; putative [Bacteriophage LL-H]	61	47	1281
509	1	554	279	gi|467484	unknown [Bacillus subtilis]	61	45	276
555	3	1916	1296	gi|141800	anthranilate synthase glutamine amidotransferase [Acinetobacter alcoaceticus]	61	42	621
569	1	1711	857	gi|467090	B2235_C2_195 [Mycobacterium leprae]	61	47	855
585	2	961	803	sp|P36686|SURE_	SURVIVAL PROTEIN SURE HOMOLOG (FRAGMENT).	61	33	159
592	3	1694	1422	gi|1221602	immunity repressor protein [Haemophilus influenzae]	61	32	273
603	1	43	357	gi|507738	Hmp [Vibrio parahaemolyticus]	61	33	315
669	1	2467	1235	gi|1146243	22.4% identity with Escherichia coli DNA-damage inducible protein ...; putative [Bacillus subtilis]	61	37	1233
675	3	805	1101	gi|403373	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis] pir|S37251|S37251 glycerophosphoryl diester phosphodiesterase - acillus subtilis	61	36	297
703	1	1656	829	gi|537181	ORF_f470 [Escherichia coli]	61	32	828
728	1	1628	816	gi|806281	DNA polymerase I [Bacillus stearothermophilus]	61	39	813
821	1	61	318	gi|709992	hypothetical protein [Bacillus subtilis]	61	38	258
856	2	2313	1567	gi|609310	portal protein gp3 [Bacteriophage HK97]	61	40	747
923	1	1081	542	gi|143213	putative [Bacillus subtilis]	61	38	540
1124	1	59	370	gi|1107541	C33D9.8 [Caenorhabditis elegans]	61	26	312
1492	1	548	276	gi|406397	unknown [Mycoplasma genitalium]	61	32	273
1602	1	46	318	gi|733522	phosphatidylinositol-4,5-diphosphate 3-kinase [Dictyostelium iscoideum]	61	34	273
2500	1	577	290	gi|1045964	hypothetical protein (GB:U1400)_297) [Mycoplasma genitalium]	61	31	288
2968	1	2	808	gi|397526	clumping factor [Staphylococcus aureus]	61	55	807
3076	1	3	248	gi|149373	ORF 1 [Lactococcus lactis]	61	41	246

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name ( )	% sim	% ident	length (nt)
3609	2	207	401	gi|1408501	homologous to N-acyl-L-amino acid amidohydrolase of Bacillus stearothermophilus [Bacillus subtilis]	61	39	195
3662	1	1477	740	gi|1303813	YqeW [Bacillus subtilis]	61	42	738
3672	1	2	442	gi|784897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir|A56390|A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	50	441
3724	1	2	220	gi|1009366	Respiratory nitrate reductase [Bacillus subtilis]	61	41	219
3728	1	3	398	gi|677943	AppD [Bacillus subtilis]	61	46	396
3884	1	3	401	gi|784897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir|A56390|A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	47	399
3971	1	3	383	gi|784897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir|A56390|A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	45	381
4038	1	661	359	gi|133995[illegible]	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	61	24	303
4041	1	546	274	gi|413953	ipa-29d gene product [Bacillus subtilis]	61	48	273
4047	1	1	402	gi|528991	unknown [Bacillus subtilis]	61	42	402
4102	1	1	345	gi|976025	HrsA [Escherichia coli]	61	46	345
4155	1	1	336	gi|784897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir|A56390|A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	50	336
4268	1	463	233	gi|450688	hsdM gene of EcoprrI gene product [Escherichia coli] pir|S38437|S38437 hsdM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	61	38	231
4374	1	542	273	gi|784897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir|A56390|A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	50	270
4389	1	2	172	gi|147516	ribokinase [Escherichia coli]	61	35	171
4621	1	2	268	gi|784897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir|A56390|A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	47	267
4663	1	27	227	gi|976025	HrsA [Escherichia coli]	61	50	201
4	6	6663	5536	gi|1408501	homologous to N-acyl-L-amino acid amidohydrolase of Bacillus stearothermophilus [Bacillus subtilis]	60	43	1128

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
11	6	3426	3725	gi|410748	ring-infested erythrocyte surface antigen (Plasmodium falciparum) pir|A25526|A25526 ring-infected erythrocyte surface antigen recursor - Plasmodium falciparum (strain FC27/Papua New Guinea) sp|P13830|RESA_PLAFF RING-INFECTED ERYTHROCYTE SURFACE ANTIGEN RE	60	24	300
11	14	11035	10313	gi|1217651	carbonyl reductase (NADPH) [Rattus norvegicus]	60	28	723
16	12	11917	12930	gi|1001453	hypothetical protein [Synechocystis sp.]	60	37	1014
33	1	26	469	gi|388109	regulatory protein [Enterococcus faecalis]	60	41	444
37	13	10914	9834	gi|1336656	Orf1 [Bacillus subtilis]	60	40	981
39	4	4364	4522	gi|4872	ORF 4 [Saccharomyces kluyveri]	60	47	159
41	1	2047	1025	gi|142822	D-alanine racemase cds [Bacillus subtilis]	60	39	1023
43	4	2474	3607	gi|468046	para-nitrobenzyl esterase [Bacillus subtilis]	60	40	1134
44	10	6756	7769	gi|414234	thiF [Escherichia coli]	60	52	1014
45	10	8874	9074	gi|343949	var1(40.0) [Saccharomyces cerevisiae]	60	44	201
56	18	27842	26430	gi|468764	mocR gene product [Rhizobium meliloti]	60	35	1413
60	2	173	388	gi|1303864	YqgQ [Bacillus subtilis]	60	33	216
63	2	357	1619	gi|467124	ureD; B229_C3_234 [Mycobacterium leprae]	60	43	1263
69	1	787	395	gi|1518853	OafA [Salmonella typhimurium]	60	36	393
88	1	1	1188	gi|1480429	putative transcriptional regulator [Bacillus stearothermophilus]	60	30	1188
92	6	4735	3881	gi|349227	transmembrane protein [Escherichia coli]	60	37	855
92	7	5996	4923	gi|466613	nikB [Escherichia coli]	60	38	1074
93	1	949	476	gi|1510925	coenzyme F420-reducing hydrogenase, beta subunit [Methanococcus jannaschii]	60	27	474
96	6	7366	7578	gi|972715	accessory protein [Carnobacterium piscicola]	60	30	213
98	6	3212	4069	gi|467425	unknown [Bacillus subtilis]	60	42	858
102	10	7158	7430	gi|143092	acetolactate synthase small subunit [Bacillus subtilis] sp|P37252|ILVN_BACSU ACETOLACTATE SYNTHASE SMALL SUBUNIT (EC .1.3.18) (AHAS) (ACETOHYDROXY-ACID SYNTHASE SMALL SUBUNIT) (ALS).	60	37	273
109	11	9127	10515	gi|1255259	o-succinylbenzoic acid (OSB) CoA ligase [Staphylococcus aureus]	60	28	1389
109	12	10499	11656	gi|141954	beta-ketothiolase [Alcaligenes eutrophus]	60	41	1158
119	2	4630	3134	gi|1524280	unknown [Mycobacterium tuberculosis]	60	45	1497

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
121	9	6957	7646	gi|1107529	ceuC gene product (Campylobacter coli)	60	35	690
140	7	7704	6013	gi|146547	kdpA (Escherichia coli)	60	45	1692
145	1	2	703	gi|1460077	unknown (Mycobacterium tuberculosis)	60	23	702
150	3	2809	2216	gi|1146230	putative (Bacillus subtilis)	60	40	594
157	2	1389	961	gi|1303975	YqjX (Bacillus subtilis)	60	30	429
158	5	5125	4769	gi|1449288	unknown (Mycobacterium tuberculosis)	60	36	357
159	1	511	257	gi|580932	murD gene product (Bacillus subtilis)	60	43	255
160	1	159	1187	gi|1204532	hypothetical protein (GB:L19201_29) (Haemophilus influenzae)	60	34	1029
161	14	8249	7866	gi|1496003	ORF3; PepY; putative oligoendopeptidase based on homology with Lactococcus lactis PepF (GenBank Accession Number Z32522) (Caldicellulosiruptor saccharolyticus)	60	34	384
172	3	1331	2110	gi|485280	28.2 kDa protein (Streptococcus pneumoniae)	60	33	780
173	2	4082	2460	gi|1524397	glycine betaine transporter OpuD (Bacillus subtilis)	60	41	1623
173	4	5463	4953	gi|1100737	NADP dependent leukotriene b4 12-hydroxydehydrogenase (Sus scrofa)	60	44	1011
198	1	3	995	gi|413943	ipa-19d gene product (Bacillus subtilis)	60	42	993
201	4	3641	4573	sp|P37028|YADT_	HYPOTHETICAL 29.4 KD PROTEIN IN HEML-PFS INTERGENIC REGION PRECURSOR.	60	37	933
203	3	3269	2415	gi|927798	D9719.34p; CAI: 0.14 (Saccharomyces cerevisiae)	60	43	855
206	9	12234	12515	sp|P37347|YECD_	HYPOTHETICAL 21.8 KD PROTEIN IN ASPS 5'REGION.	60	47	282
212	4	1213	1410	gi|332711	hemagglutinin-neuraminidase fusion protein (Human parainfluenza irus 3)	60	34	198
214	1	65	1153	gi|1204366	hypothetical protein (GB:U14003_130) (Haemophilus influenzae)	60	36	1089
237	1	2	937	gi|149377	HisD (Lactococcus lactis)	60	40	936
241	6	5696	4998	gi|1046160	hypothetical protein (GB:U00021_5) (Mycoplasma genitalium)	60	37	699
260	6	5919	6485	gi|431950	similar to a B.subtilis gene (GB: BACHEHEHY_5) (Clostridium asteurianum)	60	35	567
264	1	2432	1218	gi|397526	clumping factor (Staphylococcus aureus)	60	53	1215
267	1	3	1409	gi|148316	NaH-antiporter protein (Enterococcus hirae)	60	27	1407
275	3	3804	4595	pir|F36889|F368	leuD 3'-region hypothetical protein - Lactococcus lactis subsp. lactis (strain IL1403)	60	35	792
291	3	860	1198	gi|1208889	coded for by C. elegans cDNA yk130e12.5; contains C2H2-type zinc fingers (Caenorhabditis elegans)	60	33	339

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
307	6	3421	3176	gi|1070014	protein-dependent [Bacillus subtilis]	60	36	246
316	8	4957	5823	gi|413952	ipa-28d gene product [Bacillus subtilis]	60	41	867
328	4	2996	3484	gi|1204484	membrane-associated component, branched amino acid transport system [Haemophilus influenzae]	60	39	489
332	5	4887	4363	gi|1205449	colicin V production protein (pur regulon) [Haemophilus influenzae]	60	37	525
357	1	1062	532	gi|887842	single-stranded DNA-specific exonuclease [Escherichia coli]	60	41	531
375	2	96	362	gi|4057	adenylyl cyclase gene product [Saccharomyces kluyveri] r|JQ1145|OYBYK adenylate cyclase (EC 4.6.1.1) - yeast ccharomyces kluyveri)	60	47	267
397	1	66	416	gi|709999	Glucarate dehydratase [Bacillus subtilis]	60	37	351
409	1	2	163	gi|499700	glycogen phosphorylase [Saccharomyces cerevisiae]	60	35	162
453	4	914	1237	gi|1196899	unknown protein [Staphylococcus aureus]	60	36	324
453	7	3838	3620	sp|P12222|YCF1_	HYPOTHETICAL 226 KD PROTEIN (ORF 1901).	60	31	219
470	2	622	945	pir|S30782|S307	integrin homolog - yeast (Saccharomyces cerevisiae)	60	31	324
500	1	118	606	gi|467407	unknown [Bacillus subtilis]	60	36	489
503	3	752	982	gi|167835	myosin heavy chain [Dictyostelium discoideum]	60	34	231
505	4	2238	3563	gi|1510732	NADH oxidase [Methanococcus jannaschii]	60	26	1326
523	1	3	1043	gi|143331	alkaline phosphatase regulatory protein [Bacillus subtilis] pir|A27650|A27650 regulatory protein phoR - Bacillus subtilis sp|P23545|PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN HOR (EC 2.7.3.-).	60	41	1041
543	1	1	465	gi|1511103	cobalt transport ATP-binding protein O [Methanococcus jannaschii]	60	40	465
545	1	1	726	gi|1498192	putative [Pseudomonas aeruginosa]	60	40	726
556	1	2	1054	gi|1477402	tex gene product [Bordetella pertussis]	60	42	1053
578	1	974	489	gi|1205129	H. influenzae predicted coding region HI0882 [Haemophilus influenzae]	60	42	486
594	1	1	624	gi|1212755	adenylyl cyclase [Aeromonas hydrophila]	60	45	624
604	1	3	530	gi|145925	fecB [Escherichia coli]	60	42	528
620	1	926	465	gi|1205483	bicyclomycin resistance protein [Haemophilus influenzae]	60	33	462
630	2	871	1122	gi|1486242	unknown [Bacillus subtilis]	60	41	252
645	2	574	425	gi|1205136	serine hydroxymethyltransferase (serine methylase) [Haemophilus influenzae]	60	28	150

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
684	1	1082	843	gi|1205538	hypothetical protein (GB:U14003_302) [Haemophilus influenzae]	60	39	240
786	1	967	485	gi|1402944	orfRM1 gene product [Bacillus subtilis]	60	46	483
844	1	588	346	gi|790943	urea amidolyase [Bacillus subtilis]	60	40	243
851	1	1	726	gi|159661	GMP reductase [Ascaris lumbricoides]	60	41	726
871	1	1746	874	gi|1001493	hypothetical protein [Synechocystis sp.]	60	39.	873
896	1	1558	839	gi|604926	NADH dehydrogenase, subunit 5 [Schizophyllum commune] sp|P50368|NU5M_SCHCO NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC .6.5.3).	60	39	720
908	2	448	753	gi|662880	novel hemolytic factor [Bacillus cereus]	60	31	306
979	1	2	595	gi|1429255	putative; orf1 [Bacillus subtilis]	60	30	594
1078	1	669	502	gi|581055	inner membrane copper tolerance protein [Escherichia coli] gi|871029 disulphide isomerase like protein [Escherichia coli] pir|S47295|S47295 inner membrane copper tolerance protein - scherichia coli	60	40	168
1112	1	1150	620	gi|407885	ORF3 [Streptomyces griseus]	60	34	531
1135	1	484	275	gi|1171407	Vps8p [Saccharomyces cerevisiae]	60	36	210
1146	1	17	562	gi|1239981	hypothetical protein [Bacillus subtilis]	60	36	546
1291	1	716	360	pir|S57530|S575	carboxyl esterase - Acinetobacter calcoaceticus	60	30	357
1332	1	336	169	gi|1222056	aminotransferase [Haemophilus influenzae]	60	44	168
1429	1	3	146	gi|1205619	ferritin like protein [Haemophilus influenzae]	60	39	144
1722	1	570	286	gi|240052	dihydroflavonol-4-reductase, DFR [Hordeum vulgare=barley, cv. Gula, eptide, 354 aa]	60	36	285
2350	1	385	200	gi|497626	ORF 1 [Plasmid pAQ1]	60	20	186
2936	1	519	310	gi|508981	prephenate dehydratase [Bacillus subtilis]	60	48	210
3027	1	568	302	gi|1146199	putative [Bacillus subtilis]	60	37	267
3084	1	20	208	gi|1407784	orf-1; novel antigen [Staphylococcus aureus]	60	51	189
3155	1	2	226	gi|1046097	cytadherence-accessory protein [Mycoplasma genitalium]	60	34	225
3603	1	368	186	gi|510108	mitochondrial long-chain enoyl-CoA hydratase/3-hydroxycyl-CoA ehydrogenase alpha-subunit [Rattus norvegicus]	60	42	183
3665	1	486	244	gi|151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mevalonii] pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	60	42	243
3747	1	3	146	gi|474192	iucC gene product [Escherichia coli]	60	36	144

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3912	1	3	335	gi|488695	novel antigen; orf-2 [Staphylococcus aureus]	60	44	333
4072	1	3	272	gi|405879	yeiH [Escherichia coli]	60	33	270
4134	1	510	352	gi|780656	chemoreceptor protein [Rhizobium leguminosarum bv. viciae] gi|780656 chemoreceptor protein [Rhizobium leguminosarum bv. iciae]	60	28	159
4207	2	677	402	gi|602031	similar to trimethylamine DH [Mycoplasma capricolum] pir|S49950|S49950 probable trimethylamine dehydrogenase (EC .5.99.7) - Mycoplasma capricolum (SGC3) (fragment)	60	41	276
4243	1	127	324	gi|899317	peptide synthetase module [Microcystis aeruginosa] pir|S49111|S49111 probable amino acid activating domain - icrocystis aeruginosa (fragment) (SUB 144-528)	60	42	198
4110	1	624	313	gi|508980	pheD [Bacillus subtilis]	60	28	312
4345	1	343	173	gi|510108	mitochondrial long-chain enoyl-CoA hydratase/3-hydroxycyl-CoA ehydrogenase alpha-subunit [Rattus norvegicus]	60	42	171
4382	1	498	280	gi|47382	acyl-CoA-dehydrogenase [Streptomyces purpurascens]	60	48	219
4474	1	53	223	gi|510108	mitochondrial long-chain enoyl-CoA hydratase/3-hydroxycyl-CoA ehydrogenase alpha-subunit [Rattus norvegicus]	60	42	171
23	4	4518	3523	gi|426446	VipB protein [Salmonella typhi]	59	39	996
33	2	707	1483	pir|S48604|S486	hypothetical protein - Mycoplasma capricolum (SGC3) (fragment)	59	33	777
33	5	4651	5853	gi|6721	F59B2.3 [Caenorhabditis elegans]	59	33	1203
37	2	3328	2299	gi|142833	ORF2 [Bacillus subtilis]	59	37	930
38	21	16784	16593	gi|912576	BiP [Phaeodactylum tricornutum]	59	40	192
52	3	2648	2349	gi|536972	ORF_o30a [Escherichia coli]	59	44	300
54	12	14181	13402	gi|483940	transcription regulator [Bacillus subtilis]	59	37	780
57	3	4397	3339	gi|508176	Gat-1-P-DH, NAD dependent [Escherichia coli]	59	40	1059
66	1	986	495	gi|1303901	YqhT [Bacillus subtilis]	59	34	492
67	7	6552	7460	gi|912461	nikC [Escherichia coli]	59	37	909
70	7	5383	6366	gi|1399822	PhoD precursor [Rhizobium meliloti]	59	46	984
78	1	1	1449	gi|971345	unknown, similar to E.coli cardiolipin synthase [Bacillus subtilis] sp|P45860|YWIE_BACSU HYPOTHETICAL 58.2 PROTEIN IN NARI-ACDA NTERGENIC REGION.	59	39	1449
82	10	14329	15534	gi|490328	LORF F [unidentified]	59	44	1206

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
89	2	1602	958	gi|642801	unknown (Saccharomyces cerevisiae)	59	32	645
96	4	4940	5473	gi|1333802	protein of unknown function (Rhodobacter capsulatus)	59	33	534
98	1	2	820	gi|467421	similar to B. subtilis DnaH (Bacillus subtilis)	59	34	819
119	1	166	1557	gi|143122	ORF B; putative (Bacillus firmus)	59	36	1392
120	10	6214	6756	gi|15354	ORF 55.9 (Bacteriophage T4)	59	39	543
120	16	12476	13510	gi|1086575	BetA (Rhizobium meliloti)	59	44	1035
123	1	386	195	gi|984737	catalase (Campylobacter jejuni)	59	38	192
130	1	370	645	gi|1256634	25.8% identity over 120 aa with the Synenococcus sp. MpeV protein; putative (Bacillus subtilis)	59	31	276
131	4	5278	5712	gi|1510655	hypothetical protein (SP:P42297) (Methanococcus jannaschii)	59	39	435
164	1	3	509	gi|1001342	hypothetical protein (Synechocystis sp.)	59	41	507
164	4	1529	2821	gi|1205165	hypothetical protein (SP:P37764) (Haemophilus influenzae)	59	35	1293
164	19	19643	21376	gi|1001381	hypothetical protein (Synechocystis sp.)	59	34	1734
173	3	4727	3717	gi|1184121	auxin-induced protein (Vigna radiata)	59	50	1011
179	2	2218	1688	gi|143036	unidentified gene product (Bacillus subtilis)	59	33	531
195	12	12669	11503	gi|762778	NifS gene product (Anabaena azollae)	59	41	1167
201	5	4702	5670	gi|1510240	hemin permease (Methanococcus jannaschii)	59	32	969
201	7	5719	6315	gi|1511456	M. jannaschii predicted coding region MJ1437 (Methanococcus jannaschii)	59	34	597
209	1	102	461	gi|1204666	hypothetical protein (GB:X73124_53) (Haemophilus influenzae)	59	42	360
214	3	1050	2234	gi|551531	2-nitropropane dioxygenase (Williopsis saturnus)	59	36	1185
214	5	3293	4135	gi|1303709	YrkJ (Bacillus subtilis)	59	32	843
217	2	3381	2167	gi|290489	dfp (CG Site No. 18430) (Escherichia coli)	59	44	1215
237	5	3078	3785	gi|149382	HisA (Lactococcus lactis)	59	38	708
251	2	376	960	gi|1303791	YqeJ (Bacillus subtilis)	59	34	585
286	1	1621	812	gi|146551	transmembrane protein (kdpD) (Escherichia coli)	59	31	810
316	5	4978	3860	gi|405879	yeiH (Escherichia coli)	59	32	1119
370	3	600	761	gi|1303794	YqeM (Bacillus subtilis)	59	35	162

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% idunt	length (nt)
382	1	1009	506	gi|547513	orf3 [Haemophilus influenzae]	59	34	504
391	3	1620	1273	gi|152901	ORF 3 [Spirochaeta aurantia]	59	37	348
406	3	2805	1705	gi|709992	hypothetical protein [Bacillus subtilis]	59	34	1101
426	5	3802	3245	gi|1204610	iron(III) dicitrate transport ATP-binding protein FECE [Haemophilus influenzae]	59	36	558
429	2	1513	1148	gi|1064809	homologous to sp:HTRA_ECOLI [Bacillus subtilis]	59	42	366
460	2	708	1301	gi|466882	pps1; B1496_C2_189 [Mycobacterium leprae]	59	37	594
461	4	2212	3135	gi|1498295	homoserine kinase homolog [Streptococcus pneumoniae]	59	37	924
473	1	2929	1607	gi|147989	trigger factor [Escherichia coli]	59	40	1323
480	8	5862	6110	gi|1205311	(3R)-hydroxymyristol acyl carrier protein dehydrase [Haemophilus influenzae]	59	40	249
521	1	14	1354	pir|A25620|A256	staphylocoagulase - Staphylococcus aureus (fragment)	59	32	1341
534	4	2994	4073	gi|153746	mannitol-phosphate dehydrogenase [Streptococcus mutans] pir|C44798|C44798 mannitol-phosphate dehydrogenase MtlD - treptococcus mutans	59	36	1080
535	1	1	954	gi|1469939	group B oligopeptidase PepB [Streptococcus agalactiae]	59	33	954
551	3	2836	3186	gi|1204511	bacterioferritin comigratory protein [Haemophilus influenzae]	59	45	351
573	2	449	940	gi|386681	ORF YAL022 [Saccharomyces cerevisiae]	59	36	492
650	1	5	748	gi|396400	similar to eukaryotic Na+/H+ exchangers [Escherichia coli] sp|P32703|YJCE_ECOLI HYPOTHETICAL 60.5 KD PROTEIN IN SOXR-ACS NTERGENIC REGION (O549).	59	30	744
664	1	566	285	gi|1262748	LukF-PV like component [Staphylococcus aureus]	59	33	282
670	1	3	455	gi|1122758	unknown [Bacillus subtilis]	59	42	453
674	3	543	929	gi|293033	integrase [Bacteriophage phi-LC3]	59	46	387
758	1	349	176	gi|1500472	M. jannaschii predicted coding region MJ1577 [Methanococcus jannaschii]	59	37	174
771	2	2270	1461	gi|522150	bromoperoxidase BPO-A1 [Streptomyces aureofaciens] sp|P33912|BPA1_STRAU NON-HAEM BROMOPEROXIDASE BPO-A1 (EC 1.11.1.-) BROMIDE PEROXIDASE) (BPO1). (SUB 2-275)	59	44	810
825	1	2191	1097	gi|397526	clumping factor [Staphylococcus aureus] ,	59	47	1095
1052	2	1094	723	gi|289262	comE ORF3 [Bacillus subtilis]	59	36	372
1152	1	373	188	gi|1276668	ORF238 gene product [Porphyra purpurea]	59	37	186

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1198	1	492	247	gi|142419	ATP-dependent nuclease [Bacillus subtilis]	59	26	246
1441	1	468	235	gi|1045942	glycyl-tRNA synthetase [Mycoplasma genitalium]	59	37	234
2103	1	1	186	gi|459250	triacylglycerol lipase [Galactomyces geotrichum]	59	33	186
2205	1	793	398	gi|1303794	YqeM [Bacillus subtilis]	59	38	396
2578	1	484	284	gi|258003	insulin-like growth factor binding protein complex acid-labile ubunit [rats, liver, Peptide, 603 aa]	59	48	201
2967	2	145	348	gi|1212730	YqhK [Bacillus subtilis]	59	44	204
3012	1	3	248	gi|173571	neurofilament protein NF70 [Helix aspersa]	59	31	246
3544	1	3	401	gi|1055218	crotonase [Clostridium acetobutylicum]	59	42	399
3548	1	3	401	gi|1055218	crotonase [Clostridium acetobutylicum]	59	42	399
3580	1	698	351	gi|1055218	crotonase [Clostridium acetobutylicum]	59	42	348
3720	1	722	363	gi|1408494	homologous to penicillin acylase [Bacillus subtilis]	59	36	360
4171	1	3	296	gi|1055218	crotonase [Clostridium acetobutylicum]	59	42	294
4305	1	618	310	gi|1524193	unknown [Mycobacterium tuberculosis]	59	39	309
18	1	1242	622	gi|146913	N-acetylglucosamine transport protein [Escherichia coli] pir|B29895|WQEC2N phosphotransferase system enzyme II (EC 2.7.1.69), N-acetylglucosamine-specific - Escherichia coli sp|P09323|PTAA_ECOLI PTS SYSTEM, N-ACETYLGLUCOSAMINE-SPECIFIC IIABC OMPONENT (EIIA	58	43	621
20	7	7020	5845	gi|50502	collagen alpha chain precursor (AA -27 to 1127) [Mus musculus]	58	50	1176
21	5	3234	3626	gi|1054860	phosphoribosyl anthranilate isomerase [Thermotoga maritima]	58	32	393
23	2	2841	1669	gi|1276880	EpsG [Streptococcus thermophilus]	58	29	1173
23	10	9101	8090	pir|A31133|A311	diaminopimelate decarboxylase (EC 4.1.1.20) - Pseudomonas aeruginosa	58	37	1212
38	29	22555	22884	gi|973249	vestitone reductase [Medicago sativa]	58	37	330
44	1	2	406	gi|289272	ferrichrome-binding protein [Bacillus subtilis]	58	33	405
45	1	1	552	gi|29464	embryonic myosin heavy chain (1085 AA) [Homo sapiens] ir|S12460|S12460 myosin beta heavy chain - human	58	33	552
55	2	759	538	gi|158852	glucose regulated protein [Echinococcus multilocularis]	58	32	222
62	13	8493	8068	gi|975353	kinase-associated protein 8 [Bacillus subtilis]	58	35	426
63	3	1553	1717	gi|166926	[Arabidopsis thaliana unidentified mRNA sequence, complete cds.], one product [Arabidopsis thaliana]	58	35	165

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
67	13	12017	11229	gi|1228083	NADH dehydrogenase subunit 2 (Chorthippus parallelus)	58	41	789
96	8	8208	9167	gi|709992	hypothetical protein (Bacillus subtilis)	58	42	960
107	2	2065	1364	gi|806327	Escherichia coli hrpA gene for A protein similar to yeast PRP16 and RP22 (Escherichia coli)	58	37	702
112	7	4519	5613	gi|155588	glucose-fructose oxidoreductase (Zymomonas mobilis) pir|A42289|A42289 glucose-fructose oxidoreductase (EC 1.1.-.-) recursor - Zymomonas mobilis	58	38	1095
114	6	7318	6503	gi|1377843	unknown (Bacillus subtilis)	58	38	816
143	2	2261	1395	pir|A45605|A456	mature-parasite-infected erythrocyte surface antigen MESA - Plasmodium falciparum	58	31	867
151	2	717	950	gi|1370261	unknown (Mycobacterium tuberculosis)	58	31	234
154	6	6015	4627	gi|1209277	pCTHom1 gene product (Chlamydia trachomatis)	58	41	1389
154	16	14281	13541	gi|146613	DNA ligase (EC 6.5.1.2) (Escherichia coli)	58	39	741
155	3	2269	1892	gi|1303917	YqiB (Bacillus subtilis)	58	34	378
174	1	1056	539	gi|904198	hypothetical protein (Bacillus subtilis)	58	26	528
189	4	1533	1769	gi|467383	DNA binding protein (probable) (Bacillus subtilis)	58	25	237
201	3	2669	3307	gi|1511453	endonuclease III (Methanococcus jannaschii)	58	34	639
208	1	2	238	gi|1276729	phycobilisome linker polypeptide (Porphyra purpurea)	58	29	237
220	11	14575	13058	gi|397526	clumping factor (Staphylococcus aureus)	58	51	1518
231	3	1629	1474	gi|1002520	HutS (Bacillus subtilis)	58	45	156
233	6	4201	3497	gi|1463023	No definition line found (Caenorhabditis elegans)	58	39	705
243	10	9303	10082	gi|537207	ORF_f277 (Escherichia coli)	58	32	780
257	1	331	1143	gi|1340128	ORF1 (Staphylococcus aureus)	58	44	813
302	2	460	801	gi|40174	ORF X (Bacillus subtilis)	58	34	342
307	11	6984	6127	gi|1303842	YqfU (Bacillus subtilis)	58	30	858
321	3	1914	2747	gi|1239996	hypothetical protein (Bacillus subtilis)	58	41	834
342	4	2724	3497	gi|454838	ORF 6; putative (Pseudomonas aeruginosa)	58	41	774
348	1	1	663	gi|467478	unknown (Bacillus subtilis)	58	36	663
401	2	384	605	gi|143407	para-aminobenzoic acid synthase, component I (pab) (Bacillus ubtilis)	58	53	222

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
437	1	325	1554	gi|1303866	YqgS [Bacillus subtilis]	58	35	1230
445	1	105	1442	gi|581583	protein A [Staphylococcus aureus]	58	32	1338
453	3	789	965	gi|1009455	unknown [Schizosaccharomyces pombe]	58	34	177
453	5	2748	2047	gi|537214	yjjG gene product [Escherichia coli]	58	40	702
479	2	731	1444	gi|1256621	26.7% of identity in 165 aa to a Thermophilic bacterium hypothetical protein 6; putative [Bacillus subtilis]	58	36	714
490	1	909	547	gi|580920	rodD (gtaA) polypeptide (AA 1-673) [Bacillus subtilis] pir|S06048|S06048 probable rodD protein - Bacillus subtilis sp|P13484|TAGE_BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) LPHA-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHOIC ACID BIOSYNTHESIS ROTEIN E).	58	36	363
517	1	1	1164	sp|P47264|Y018_	HYPOTHETICAL HELICASE MG018.	58	30	1164
517	6	4182	4544	gi|453422	orf268 gene product [Mycoplasma hominis]	58	29	363
546	3	2802	4019	gi|886052	restriction modification system S subunit [Spiroplasma citri] gi|886052 restriction modification system S subunit [Spiroplasma itri]	58	37	1218
562	1	3	179	gi|43831	nifS protein (AA 1-400) [Klebsiella pneumoniae]	58	34	177
600	2	1347	1156	gi|1183839	unknown [Pseudomonas aeruginosa]	58	48	192
604	2	1231	1001	gi|1001353	hypothetical protein [Synechocystis sp.]	58	41	231
619	1	1	504	gi|903748	integral membrane protein [Homo sapiens]	58	43	504
625	1	2	364	gi|1208474	hypothetical protein [Synechocystis sp.]	58	43	363
635	1	1492	755	gi|1510995	transaldolase [Methanococcus jannaschii]	58	41	738
645	1	1	846	gi|677882	ileal sodium-dependent bile acid transporter [Rattus norvegicus] gi|677882 ileal sodium-dependent bile acid transporter [Rattus orvegicus]	58	33	846
645	3	906	1556	gi|1239999	hypothetical protein [Bacillus subtilis]	58	41	651
665	1	771	532	gi|1204262	hypothetical protein (GB:L10328_61) [Haemophilus influenzae]	58	39	240
674	1	635	327	gi|498817	ORF8; homologous to small subunit of phage terminases [Bacillus ubtilis]	58	39	309
675	2	1312	806	gi|42181	osmC gene product [Escherichia coli]	58	28	507
745	1	618	310	gi|1205432	coenzyme PQQ synthesis protein III (pqqIII) [Haemophilus influenzae]	58	32	309
799	2	242	1174	gi|1204669	collagenase [Haemophilus influenzae]	58	36	933
800	2	1096	614	gi|171963	tRNA isopentenyl transferase [Saccharomyces cerevisiae] sp|P07884|MOD5_YEAST TRNA ISOPENTENYLTRANSFERASE (EC 2.5.1.8) ISOPENTENYL-DIPHOSPHATE: TRNA ISOPENTENYLTRANSFERASE) (IPP RANSFERASE) (IPPT).	58	37	483

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
854	1	1108	605	gi|466778	lysine specific permease (Escherichia coli)	58	44	504
885	1	481	242	gi|861199	protoporphyrin IX Mg-chelatase subunit precursor (Hordeum vulgare)	58	33	240
891	1	3	527	gi|1293660	AbsA2 (Streptomyces coelicolor)	58	31	525
942	1	931	467	gi|405567	traH (Plasmid pSK41)	58	30	465
1002	1	952	521	gi|577649	proLUKM (Staphylococcus aureus)	58	34	432
1438	1	1	261	gi|581558	isoleucyl tRNA synthetase (Staphylococcus aureus) sp|P41368|SYIP_STAAU ISOLEUCYL-TRNA SYNTHETASE, MUPIROCIN RESISTANT EC 6.1.1.5) (ISOLEUCINE-- TRNA LIGASE) (ILERS) (MUPIROCIN RESISTANCE ROTEIN).	58	30	261
1442	1	2	463	gi|971394	similar to Acc.No. D26185 (Escherichia coli)	58	34	462
1873	1	480	241	gi|1339951	small subunit of NADH-dependent glutamate synthase (Plectonema boryanum)	58	38	240
1876	1	3	158	gi|529216	No definition line found (Caenorhabditis elegans) sp|P46503|YLX7_CAEEL HYPOTHETICAL 7.3 KD PROTEIN F23F12.7 IN HROMOSOME III.	58	33	156
1989	1	108	401	gi|1405458	YneR (Bacillus subtilis)	58	29	294
2109	1	3	401	gi|1001801	hypothetical protein (Synechocystis sp.)	58	31	399
2473	1	288	145	gi|510140	ligoendopeptidase F (Lactococcus lactis)	58	38	144
2523	1	452	228	gi|644873	catabolic dehydroquinate dehydratase (Acinetobacter calcoaceticus)	58	37	225
3041	1	2	211	gi|1205367	oligopeptide transport ATP-binding protein (Haemophilus influenzae)	58	39	210
3094	1	3	263	gi|1185288	isochorismate synthase (Bacillus subtilis)	58	38	261
3706	1	3	383	gi|456614	mevalonate kinase (Arabidopsis thaliana)	58	48	381
3854	1	1	402	gi|808869	human gcp372 (Homo sapiens)	58	32	402
4082	1	51	224	gi|508551	ribulose-1,5 bisphosphate carboxylase large subunit -methyltransferase (Pisum sativum)	58	37	174
4278	1	3	206	gi|180189	cerebellar-degeneration-related antigen (CDR34) (Homo sapiens) gi|182737 cerebellar degeneration-associated protein (Homo sapiens) pir|A29770|A29770 cerebellar degeneration-related protein - human	58	37	204
19	7	7818	7363	gi|1001516	hypothetical protein (Synechocystis sp.)	57	31	456
23	11	9663	8872	gi|606066	ORF_f256 (Escherichia coli)	57	29	792
31	1	4801	2402	gi|153146	ORF3 (Streptomyces coelicolor)	57	32	2400
38	14	11611	10796	gi|144859	ORF B (Clostridium perfringens)	57	31	816
46	14	12063	13046	gi|1001319	hypothetical protein (Synechocystis sp.)	57	25	984

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
51	3	1411	1187	pir|B33856|B338	hypothetical 80K protein - Bacillus sphaericus	57	38	225
54	1	1	453	gi|684950	staphylococcal accessory regulator A (Staphylococcus aureus)	57	31	453
75	1	3	239	gi|1000470	C27B7.7 (Caenorhabditis elegans)	57	42	237
92	5	3855	3061	gi|143607	sporulation protein (Bacillus subtilis)	57	35	795
96	3	4006	4773	gi|144297	acetyl esterase (XynC) (Caldocellum saccharolyticum) pir|B37202|B37202 acetylesterase (EC 3.1.1.6) (XynC) - Caldocellum saccharolyticum	57	34	768
107	3	1480	2076	gi|460955	TagE (Vibrio cholerae)	57	42	597
109	8	5340	5933	gi|1438846	Unknown (Bacillus subtilis)	57	41	594
112	9	6679	7701	gi|1486250	unknown (Bacillus subtilis)	57	33	1023
114	4	6384	4108	gi|871456	putative alpha subunit of formate dehydrogenase (Methanobacterium thermoautotrophicum)	57	37	2277
126	2	430	1053	gi|288301	ORF2 gene product (Bacillus megaterium)	57	37	624
131	5	6537	6277	gi|1511160	M. jannaschii predicted coding region MJ1163 (Methanococcus jannaschii)	57	38	261
133	3	2668	2201	gi|1303912	YqhW (Bacillus subtilis)	57	40	468
133	4	3383	2784	gi|1221884	(urea?) amidolyase (Haemophilus influenzae)	57	37	600
147	4	2164	1694	gi|467469	unknown (Bacillus subtilis)	57	33	471
160	2	1293	1060	gi|558604	chitin synthase 2 (Neurospora crassa)	57	28	234
163	8	5687	4764	gi|145580	rarD gene product (Escherichia coli)	57	38	924
168	6	4336	5325	gi|39782	33kDa lipoprotein (Bacillus subtilis)	57	32	990
170	5	3297	3455	gi|603404	Yer164p (Saccharomyces cerevisiae)	57	37	159
221	6	8026	6809	gi|1136221	carboxypeptidase (Sulfolobus solfataricus)	57	32	1218
228	3	1348	1791	gi|288969	fibronectin binding protein (Streptococcus dysgalactiae) pir|S33850|S33850 fibronectin-binding protein - Streptococcus ysgalactiae	57	32	444
263	4	4411	3686	gi|1185002	dihydrodipicolinate reductase (Pseudomonas syringae pv. tabaci)	57	42	726
276	1	494	255	gi|396380	No definition line found (Escherichia coli)	57	40	240
283	2	335	1324	gi|773349	BirA protein (Bacillus subtilis)	57	32	990
297	1	469	236	gi|1334820	reading frame V (Cauliflower mosaic virus)	57	46	234
342	3	1993	2805	gi|1204431	hypothetical protein (SP:P33644) (Haemophilus influenzae)	57	35	813

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
375	6	3340	3741	gi|385177	cell division protein [Bacillus subtilis]	57	26	402
433	6	3286	4011	gi|1524117	alpha-acetolactate decarboxylase [Lactococcus lactis]	57	40	726
470	3	903	1145	gi|404819	protein serine/threonine kinase [Toxoplasma gondii]	57	30	243
487	5	1391	1723	gi|507323	ORF1 [Bacillus stearothermophilus]	57	28	333
498	1	274	852	gi|1334549	NADH-ubiquinone oxidoreductase subunit 4L [Podospora anserina]	57	34	579
503	1	343	173	gi|1502283	organic cation transporter OCT2 [Rattus norvegicus]	57	30	171
505	2	1619	1284	gi|466884	B1496_C2_191 [Mycobacterium leprae]	57	40	336
519	2	1182	2549	gi|1303707	YrkH [Bacillus subtilis]	57	34	1368
522	2	3234	1945	gi|1064809	homologous to sp:HCRA_ECOLI [Bacillus subtilis]	57	36	1290
538	2	909	1415	gi|153179	phosphorinothyricin N-acetyltransferase [Streptomyces coelicolor] pir|JH0246|JH0246 phosphinothricin N-acetyltransferase (EC 2.3.1.-) Streptomyces coelicolor	57	40	507
547	1	968	486	gi|467340	unknown [Bacillus subtilis]	57	50	483
599	1	1062	532	sp|P20692|PYRA_	PHEPHENATE DEHYDROGENASE (EC 1.3.1.12) (PDH).	57	41	531
620	2	757	572	gi|1107894	unknown [Schizosaccharomyces pombe]	57	38	186
622	2	1600	1130	gi|173028	thioredoxin II [Saccharomyces cerevisiae]	57	39	471
625	2	362	1114	gi|1262366	hypothetical protein [Mycobacterium leprae]	57	34	753
680	1	1	204	gi|143544	RNA polymerase sigma-30 factor [Bacillus subtilis] pir|A28625|A28625 transcription initiation factor sigma H - acillus subtilis	57	30	204
690	1	3	629	gi|466520	pocR [Salmonella typhimurium]	57	29	627
696	1	2	433	gi|413972	ipa-48r gene product [Bacillus subtilis]	57	33	432
704	1	36	638	gi|1499931	M. jannaschii predicted coding region MJ1083 [Methanococcus jannaschii]	57	36	603
732	1	2316	1621	gi|1418999	orf4 [Lactobacillus sake]	57	37	696
746	1	451	227	gi|392973	Rab3 [Aplysia californica]	57	42	225
757	1	20	466	gi|43979	L.curvatus small cryptic plasmid gene for rep protein [Lactobacillus rvatus]	57	45	447
862	1	2	295	gi|1303827	YqfI [Bacillus subtilis]	57	21	294
1049	1	907	455	gi|1510108	ORF-1 [Agrobacterium tumefaciens]	57	35	453
1117	1	1387	695	gi|896286	NH2 terminus uncertain [Leishmania tarentolae]	57	28	693

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name ( )	% sim	% ident	length (nt)
1136	1	2	322	gi|1303853	YqqF (Bacillus subtilis)	57	38	321
1144	2	1033	611	gi|310083	voltage-activated calcium channel alpha-1 subunit (Rattus orvegicus)	57	46	423
1172	1	1472	738	gi|1511146	H. jannaschii predicted coding region MJ1143 (Methanococcus jannaschii)	57	28	735
1500	2	746	558	gi|142780	putative membrane protein; putative (Bacillus subtilis)	57	35	189
1676	1	659	399	gi|313777	uracil permease (Escherichia coli)	57	31	261
2481	1	2	400	gi|1237015	ORF4 (Bacillus subtilis)	57	23	399
3099	1	3	230	gi|1204540	isochorismate synthase (Haemophilus influenzae)	57	19	228
3122	1	360	181	gi|882472	ORF_o464 (Escherichia coli)	57	40	180
3560	1	2	361	gi|153490	tetracenomycin C resistance and export protein (Streptomyces laucescens)	57	37	360
3850	1	856	434	gi|155588	glucose-fructose oxidoreductase (Zymomonas mobilis) pir|A42289|A42289 glucose-fructose oxidoreductase (EC 1.1.-.-) recursor - Zymomonas mobilis	57	40	423
3931	1	704	354	gi|413953	ipa-29d gene product (Bacillus subtilis)	57	36	351
3993	1	1	384	gi|151259	HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas mevalonii) pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	57	39	384
4065	1	793	398	pir|JV0037|RDEC	nitrate reductase (EC 1.7.99.4) alpha chain - Escherichia coli	57	31	396
4100	1	596	300	gi|1086633	T06C10.5 gene product (Caenorhabditis elegans)	57	47	297
4163	1	571	287	gi|21512	patatin (Solanum tuberosum)	57	50	285
4267	2	631	335	gi|1000365	SpoIIIAG (Bacillus subtilis)	57	38	297
4358	1	3	302	gi|298032	EF (Streptococcus suis)	57	32	300
4389	2	108	290	gi|405894	1-phosphofructokinase (Escherichia coli)	57	37	183
4399	1	2	232	gi|1483603	Pristinamycin I synthase I (Streptomyces pristinaespiralis)	57	35	231
4481	1	572	288	gi|405879	yeiH (Escherichia coli)	57	44	285
4486	1	512	258	gi|515938	glutamate synthase (ferredoxin) (Synechocystis sp.) pir|S46957|S46957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - ynechocystis sp.	57	42	255
4510	1	481	242	gi|1205301	leukotoxin secretion ATP-binding protein (Haemophilus influenzae)	57	38	240
4617	1	468	256	gi|1511222	restriction modification enzyme, subunit M1 (Methanococcus jannaschii)	57	35	213
4	11	12201	11524	gi|149204	histidine utilization repressor G (Klebsiella aerogenes) pir|A36730|A36730 hutG protein - Klebsiella pneumoniae (fragment) sp|P19452|HUTG_KLEAE FORMIMINOGLUTAMASE (EC 3.5.3.8) FORMIMINOGLUTAMATE HYDROLASE) (HISTIDINE UTILIZATION PROTEIN G) FRAGMENT).	56	31	678

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name ( )	% sim	% ident	length (nt)
22	8	4248	5177	gi|1322222	RACH1 [Homo sapiens]	56	33	930
38	28	21179	22264	gi|1480705	lipoate-protein ligase [Mycoplasma capricolum]	56	34	1086
44	3	1861	2421	gi|490320	Y gene product (unidentified)	56	31	561
44	15	10103	10606	gi|1205099	hypothetical protein (GB:L19201_1) [Haemophilus influenzae]	56	39	504
50	6	4820	5161	gi|209931	fiber protein [Human adenovirus type 5]	56	48	342
53	4	2076	2972	gi|623476	transcriptional activator [Providencia stuartii] sp|P43463|AARP_PROST TRANSCRIPTIONAL ACTIVATOR AARP.	56	30	897
67	6	5656	6594	gi|466613	nikB [Escherichia coli]	56	32	939
89	3	2364	1810	gi|482922	protein with homology to pail repressor of B.subtilis [Lactobacillus elbrueckii]	56	39	555
96	1	203	913	gi|145594	cAMP receptor protein (crp) [Escherichia coli]	56	35	711
109	21	18250	17846	gi|1204367	hypothetical protein (GB:U14003_278) [Haemophilus influenzae]	56	27	405
112	8	5611	6678	gi|155588	glucose-fructose oxidoreductase [Zymomonas mobilis] pir|A42289|A42289 glucose-fructose oxidoreductase (EC 1.1.-.-) recursor - Zymomonas mobilis	56	40	1068
131	3	6404	5100	gi|619724	MgtE [Bacillus firmus]	56	30	1305
138	2	65	232	gi|413948	ipa-24d gene product [Bacillus subtilis]	56	31	168
138	4	823	1521	gi|580868	ipa-22r gene product [Bacillus subtilis]	56	31	699
146	2	740	447	gi|1046009	M. genitalium predicted coding region MG309 [Mycoplasma genitalium]	56	37	294
149	2	1619	1067	gi|945380	terminase small subunit [Bacteriophage LL-H]	56	35	573
163	1	2	223	gi|143947	glutamine synthetase [Bacteroides fragilis]	56	30	222
166	5	6745	6449	gi|405792	ORF154 [Pseudomonas putida]	56	26	297
187	1	31	393	gi|311237	H(+)-transporting ATP synthase [Zea mays]	56	30	363
190	1	2	373	gi|1109686	ProX [Bacillus subtilis]	56	35	372
191	8	11538	9943	gi|581070	acyl coenzyme A synthetase [Escherichia coli]	56	35	1596
195	1	1291	647	gi|1510242	collagenase [Methanococcus jannaschii]	56	34	645
230	3	2323	2072	gi|40363	heat shock protein [Clostridium acetobutylicum]	56	39	252
238	5	3383	3775	gi|1477533	sarA [Staphylococcus aureus]	56	31	393
270	2	813	1712	gi|765073	autolysin [Staphylococcus aureus]	56	41	900

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name ( )	% sim	% ident	length (nt)
290	1	3221	1632	gi|547513	orf3 [Haemophilus influenzae]	56	34	1590
297	5	1140	1373	gi|1511556	M. jannaschii predicted coding region MJ1561 [Methanococcus jannaschii]	56	40	234
321	2	2947	1799	gi|1001801	hypothetical protein [Synechocystis sp.]	56	31	1149
359	2	1279	641	gi|46336	nolI gene product [Rhizobium meliloti]	56	26	639
371	2	360	1823	gi|145304	L-ribulokinase [Escherichia coli]	56	39	1464
391	4	1762	2409	gi|1001634	hypothetical protein [Synechocystis sp.]	56	34	648
402	1	380	192	gi|1438904	5-HT4L receptor [Homo sapiens]	56	48	189
416	4	2480	2109	gi|1408486	HS74A gene product [Bacillus subtilis]	56	31	372
424	3	1756	2334	gi|142471	acetolactate decarboxylase [Bacillus subtilis]	56	32	579
457	1	1907	1017	gi|1205194	formamidopyrimidine-DNA glycosylase [Haemophilus influenzae]	56	36	891
458	2	2423	1812	gi|15466	terminase [Bacteriophage SPP1]	56	37	612
504	2	2152	1283	gi|1142681	Lpp38 [Pasteurella haemolytica]	56	38	870
511	1	1	1284	gi|217049	brnQ protein [Salmonella typhimurium]	56	37	1284
604	3	1099	1701	gi|467109	rim; 30S Ribosomal protein S18 alanine acetyltransferase; 229_C1_170 [Mycobacterium leprae]	56	43	603
660	5	3547	3774	gi|1229106	ZK930.1 [Caenorhabditis elegans]	56	30	228
707	1	35	400	gi|153929	NADPH-sulfite reductase flavoprotein component [Salmonella typhimurium]	56	38	366
709	2	1385	1095	gi|1510801	hydrogenase accessory protein [Methanococcus jannaschii]	56	38	291
718	1	1	495	gi|413948	ipa-24d gene product [Bacillus subtilis]	56	35	495
744	1	87	677	gi|928836	repressor protein [Lactococcus lactis phage BK5-T]	56	35	591
790	1	776	399	gi|1511513	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	56	33	378
795	1	3	407	gi|1205382	cell division protein [Haemophilus influenzae]	56	34	405
813	1	19	930	gi|1222161	permease [Haemophilus influenzae]	56	28	912
855	1	3	515	gi|1256621	26.7% of identity in 165 aa to a Thermophilic bacterium hypothetical protein 6; putative [Bacillus subtilis]	56	33	513
968	1	2	466	gi|547513	orf3 [Haemophilus influenzae]	56	37	465
973	2	1049	732	gi|886022	MexR [Pseudomonas aeruginosa]	56	31	318
1203	1	5	223	gi|184251	HHG-1 [Homo sapiens]	56	34	219

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
1976	1	452	237	gi|9806	lysine-rich aspartic acid-rich protein [Plasmodium chabaudi] r|S22183|S22183 lysine/aspartic acid-rich protein - Plasmodium baudi	56	33	216
2161	1	2	400	gi|1237015	ORF4 [Bacillus subtilis]	56	27	399
2958	1	362	183	gi|466685	No definition line found [Escherichia coli]	56	26	180
2979	1	421	212	gi|1204354	spore germination and vegetative growth protein [Haemophilus influenzae]	56	40	210
2994	2	526	326	gi|836646	phosphoribosylformimino-praic ketoisomerase [Rhodobacter phaeroides]	56	29	201
3026	1	179	328	gi|143306	penicillin V amidase [Bacillus sphaericus]	56	30	150
3189	1	289	146	gi|1166604	Similar to aldehyde dehydrogenase [Caenorhabditis elegans]	56	37	144
3770	1	63	401	gi|1129145	acetyl-CoA C-acyltransferase [Mangifera indica]	56	43	339
4054	2	720	361	gi|1205355	Na+/H+ antiport [Haemophilus influenzae]	56	31	360
4145	1	1	324	gi|726095	long-chain acyl-CoA dehydrogenase [Mus musculus]	56	36	324
4200	1	505	254	gi|155588	glucose-fructose oxidoreductase (Zymomonas mobilis) pir|A42289|A42289 glucose-fructose oxidoreductase (EC 1.1.-.-) recursor - Zymomonas mobilis	56	40	252
4273	1	675	355	gi|308861	GTG start codon [Lactococcus lactis]	56	33	321
1	3	4095	3436	gi|5341	Putative orf (ORF, len:192 [Saccharomyces cerevisiae] r|S53591|S53591 hypothetical protein - yeast (Saccharomyces cevisiae)	55	25	660
11	12	9377	8505	gi|216773	haloacetate dehalogenase H-1 [Moraxella sp.]	55	32	873
12	4	5133	4534	gi|467337	unknown [Bacillus subtilis]	55	26	600
19	5	5404	5844	gi|1001719	hypothetical protein [Synechocystis sp.]	55	25	441
23	13	14087	12339	gi|474190	iucA gene product [Escherichia coli]	55	30	1749
32	7	5368	6888	gi|1340096	unknown [Mycobacterium tuberculosis]	55	37	1521
34	3	2569	1808	gi|1303968	YqjQ [Bacillus subtilis]	55	39	762
34	5	3960	3412	gi|1303962	YqjK [Bacillus subtilis]	55	33	549
36	1	1291	647	gi|606045	ORF_o118 [Escherichia coli]	55	27	645
36	6	6220	5243	gi|1001341	hypothetical protein [Synechocystis sp.]	55	31	978
47	3	3054	3821	gi|1001819	hypothetical protein [Synechocystis sp.]	55	21	768
49	1	2065	1127	gi|403373	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis] pir|S37251|S37251 glycerophosphoryl diester phosphodiesterase - acillus subtilis	55	36	939

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
67	11	8966	9565	gi|153053	norA1199 protein [Staphylococcus aureus]	55	23	600
75	3	881	1273	gi|41698	L-histidinol: NAD+ oxidoreductase (EC 1.1.1.23) (aa 1-434) scherichia coli]	55	33	393
82	9	15387	14194	gi|1136221	carboxypeptidase [Sulfolobus solfataricus]	55	35	1194
87	4	3517	4917	gi|1064812	function unknown [Bacillus subtilis]	55	26	1401
88	2	1172	1636	gi|882463	protein-N(pi)-phosphohistidine-sugar phosphotransferase [Escherichia oli]	55	35	465
92.	1	127	516	gi|1377832	unknown [Bacillus subtilis]	55	36	390
100	2	836	2035	gi|1370274	zeaxanthin epoxidase [Nicotiana plumbaginifolia]	55	36	1200
100	5	5137	4658	gi|396660	unknown open reading frame [Buchnera aphidicola]	55	29	480
108	3	4266	2986	gi|1499866	M. jannaschii predicted coding region MJ1024 [Methanococcus jannaschii]	55	31	1281
114	3	2616	1834	gi|1511367	formate dehydrogenase, alpha subunit [Methanococcus jannaschii]	55	29	783
144	3	1805	1476	gi|1100787	unkown [Saccharomyces cerevisiae]	55	35	330
165	5	6212	5508	gi|1045884	M. genitalium predicted coding region MG199 [Mycoplasma genitalium]	55	27	705
189	5	2205	2576	gi|142569	ATP synthase A subunit [Bacillus firmus]	55	35	372
191	6	9136	6857	gi|559411	B0272.3 [Caenorhabditis elegans]	55	39	2280
194	2	364	636	gi|1145768	K7 kinesin-like protein [Dictyostelium discoideum]	55	34	273
209	4	1335	1676	gi|473357	thi4 gene product [Schizosaccharomyces pombe]	55	35	342
211	2	1693	1145	gi|410130	ORFX6 [Bacillus subtilis]	55	37	549
213	2	644	1372	gi|633692	TrsA [Yersinia enterocolitica]	55	28	729
214	7	4144	5481	gi|1001793	hypothetical protein [Synechocystis sp.]	55	30	1338
221	7	11473	9197	gi|466520	pocR [Salmonella typhimurium]	55	32	2277
233	8	5908	4817	gi|1237063	unknown [Mycobacterium tuberculosis]	55	38	1092
236	4	1375	2340	gi|1146199	putative [Bacillus subtilis]	55	32	966
243	2	380	1885	gi|459907	mercuric reductase [Plasmid pI258]	55	29	1506
258	1	786	394	gi|455006	orf6 [Rhodococcus fascians]	55	36	393
281	1	126	938	gi|1408493	homologous to sp:P39127:YIDA_ECOLI hypothetical protein [Bacillus subtilis]	55	35	813
316	3	1323	2102	gi|1486447	LuxA homologue [Rhizobium sp.]	55	30	780
326	5	2968	2744	gi|1296824	proline iminopeptidase [Lactobacillus helveticus]	55	36	225

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
351	2	2322	1429	gi|1204820	hydrogen peroxide-inducible activator (Haemophilus influenzae)	55	28	894
353	4	2197	2412	gi|1272475	chitin synthase (Emericella nidulans)	55	50	216
380	1	14	379	gi|142554	ATP synthase I subunit (Bacillus megaterium)	55	37	366
383	1	462	232	gi|189272	ferrichrome-binding protein (Bacillus subtilis)	55	36	231
386	1	3	938	gi|1510251	DNA helicase, putative (Methanococcus jannaschii)	55	30	936
410	2	1208	1891	gi|1205144	multidrug resistance protein (Haemophilus influenzae)	55	27	684
483	2	411	833	gi|413934	ipa-10r gene product (Bacillus subtilis)	55	26	423
529	3	1777	1433	gi|606150	ORF_f309 (Escherichia coli)	55	33	345
555	1	1088	585	gi|143407	para-aminobenzoic acid synthase, component I (pab) (Bacillus ubtilis)	55	28	504
565	1	402	202	gi|1223961	CDP-tyvelose epimerase (Yersinia pseudotuberculosis)	55	41	201
582	1	751	452	gi|1256643	20.2% identity with NADH dehydrogenase of the Leishmania major mitochondrion, putative (Bacillus subtilis)	55	36	300
645	5	2260	2057	gi|210824	fusion protein F (Bovine respiratory syncytial virus) pir|JQ1481|VGNZBA fusion glycoprotein precursor - bovine espiratory syncytial virus (strain A51908)	55	25	204
672	2	957	2216	gi|1511333	M. jannaschii predicted coding region MJ1322 (Methanococcus jannaschii)	55	36	1260
730	1	955	479	gi|537007	ORF_f379 (Escherichia coli)	55	30	477
737	1	1859	945	gi|536963	CG Site No. 15166 (Escherichia coli)	55	30	915
742	2	228	572	gi|304160	product unknown (Bacillus subtilis)	55	38	345
817	2	1211	903	gi|1136289	histidine kinase A (Dictyostelium discoideum)	55	29	309
819	1	582	355	gi|558073	polymorphic antigen (Plasmodium falciparum)	55	22	228
832	2	1152	724	gi|40367	ORFC (Clostridium acetobutylicum)	55	32	429
840	1	769	386	gi|1205875	pseudouridylate synthase I (Haemophilus influenzae)	55	39	384
1021	1	23	529	gi|48563	beta-lactamase (Yersinia enterocolitica)	55	38	507
1026	1	60	335	gi|47804	Opp C (AA1-301) (Salmonella typhimurium)	55	26	276
1525	1	1	282	gi|1477533	sarA (Staphylococcus aureus)	55	29	282
1814	2	224	985	gi|1046078	M. genitalium predicted coding region MG369 (Mycoplasma genitalium)	55	38	762
3254	1	427	254	gi|413968	ipa-44d gene product (Bacillus subtilis)	55	30	174

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
3695	1	686	345	gi|216773	haloacetate dehalogenase H-1 [Moraxella sp.]	55	32	342
3721	1	1	312	gi|42029	ORF1 gene product [Escherichia coli]	55	31	312
3799	1	3	272	gi|42029	ORF1 gene product [Escherichia coli]	55	38	270
3889	1	22	423	gi|1129145	acetyl-CoA C-acyltransferase [Mangifera indica]	55	45	402
3916	1	2	385	gi|529754	speC [Streptococcus pyogenes]	55	38	384
3945	1	4	198	gi|476252	phase 1 flagellin [Salmonella enterica]	55	36	195
4074	1	488	246	gi|42029	ORF1 gene product [Escherichia coli]	55	38	243
4184	1	2	343	gi|1524267	unknown [Mycobacterium tuberculosis]	55	28	342
4284	1	14	208	gi|1100774	ferredoxin-dependent glutamate synthase [Synechocystis sp.]	55	36	195
4457	2	644	378	gi|180189	cerebellar-degeneration-related antigen (CDR34) [Homo sapiens] gi|182737 | cerebellar degeneration-associated protein [Homo sapiens] | pir|A29770|A2 ...0 cerebellar degeneration-related protein - human	55	38	267
4514	1	2	244	gi|216773	haloacetate dehalogenase H-1 [Moraxella sp.]	55	32	243
4599	1	432	217	gi|1129145	acetyl-CoA C-acyltransferase [Mangifera indica]	55	42	216
4606	1	416	210	gi|386120	myosin alpha heavy chain (S2 subfragment) [rabbits, masseter, eptide | Partial, 234 aa]	55	27	207
5	8	5348	4932	gi|536069	ORF YBL047c [Saccharomyces cerevisiae]	54	27	417
12	7	7166	6165	gi|1205504	homoserine acetyltransferase [Haemophilus influenzae]	54	30	1002
23	16	17086	15326	gi|474192	iucC gene product [Escherichia coli]	54	31	1761
35	1	2	979	gi|48054	small subunit ...ble Ni hydrogenase (AA 1-384) [Synechococcus sp.] | ir|S06919|H2 ...luble hydrogenase (EC 1.12.-.-) small chain - | nechococcus ...(EC ...)]	54	36	978
37	11	9437	8667	gi|537207	ORF_f277 [E... coli]	54	38	771
37	12	8165	8332	gi|1160967	palmitoyl-pr... hydrolase [Homo sapiens]	54	37	168
46	15	13025	13804	gi|438473	protein is h... with homology to E. coli ProW; putative Bacillus | subtilis]	54	28	780
56	2	203	736	gi|1256139	YbbJ [Bacillus ...]	54	34	534
57	13	11117	10179	gi|1151248	inosine-uridine preferring nucleoside hydrolase [Crithidia fasciculata]	54	32	939
66	2	516	1133	gi|1335781	Cap [Drosophila melanogaster]	54	29	618
70	10	8116	8646	gi|1399823	PhoE [Rhizob...]	54	31	531

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
70	15	12556	11801	sp|P02983|TCR_S	TETRACYCLINE RESISTANCE PROTEIN.	54	29	756
87	5	4915	5706	gi|1064811	function unknown (Bacillus subtilis)	54	33	792
92	4	3005	2289	gi|1205366	oligopeptide transport ATP-binding protein (Haemophilus influenzae)	54	33	717
103	2	2596	1556	gi|710495	protein kinase (Bacillus brevis)	54	33	1041
105	2	3585	2095	gi|143727	putative (Bacillus subtilis)	54	30	1491
112	4	2337	2732	gi|153724	HalC (Streptococcus pneumoniae)	54	41	396
127	2	1720	2493	gi|144297	acetyl esterase (XynC) (Caldocellum saccharolyticum) pir|B37202|B37202 acetylesterase (EC 3.1.1.6) (XynC) - Caldocellum saccharolyticum	54	34	774
138	5	1600	3306	gi|42473	pyruvate oxidase (Escherichia coli)	54	36	1707
152	2	525	1172	gi|1377834	unknown (Bacillus subtilis)	54	23	648
161	9	4831	5469	gi|903305	ORF73 (Bacillus subtilis)	54	28	639
161	13	6694	7251	gi|1511039	phosphate transport system regulatory protein (Methanococcus jannaschii)	54	32	558
164	6	3263	4543	gi|1204976	prolyl-tRNA synthetase (Haemophilus influenzae)	54	34	1281
164	20	21602	22243	gi|143582	spoIIIEA protein (Bacillus subtilis)	54	32	642
171	6	5683	4250	gi|436965	(malA) gene product (Bacillus stearothermophilus) pir|S43914|S43914 hypothetical protein 1 - Bacillus tearothermophilus	54	37	1434
206	18	19208	19720	gi|1240016	R09E10.3 (Caenorhabditis elegans)	54	38	513
218	2	1090	1905	gi|467378	unknown (Bacillus subtilis)	54	26	816
220	1	1322	663	gi|1353761	myosin II heavy chain (Dictyostelium fowleri)	54	22	660
220	13	12655	13059	pir|S00485|S004	gene 11-1 protein precursor - Plasmodium falciparum (fragments)	54	35	405
221	3	2030	3709	gi|1303813	YqeW (Bacillus subtilis)	54	34	1680
272	7	5055	4219	gi|62964	arylamine N-acetyltransferase (AA 1-290) (Gallus gallus) ir|S06652|XYCHY3 arylamine N-acetyltransferase (EC 2.3.1.5) (clone NAT-3) - chicken	54	33	837
316	7	4141	4701	gi|682769	mccE gene product (Escherichia coli)	54	31	561
316	10	6994	8742	gi|413951	ipa-27d gene product (Bacillus subtilis)	54	28	1749
338	3	3377	2214	gi|490328	LORF F (unidentified)	54	28	1164
341	4	3201	3614	gi|171959	myosin-like protein (Saccharomyces cerevisiae)	54	25	414

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
346	1	1820	912	gi|396400	similar to e ...aryotic Na+/H+ exchangers [Escherichia coli] sp|P32703|YJ E ...OLI HYPOTHETICAL 60.5 KD PROTEIN IN SOXR-ACS NTERGENIC REGION (O54...	54	34	909
348	2	623	1351	gi|537109	ORF_f343a [E. ...coli]	54	34	729
378	2	1007	1942	sp|P02983|TCR_S	TETRACYCLINE R... ISTANCE PROTEIN.	54	31	936
408	6	4351	5301	gi|474190	iucA gene pro...t [Escherichia coli]	54	29	951
444	9	7934	8854	gi|216267	ORF2 [Bacillus ...egaterium]	54	32	921
463	2	2717	2229	gi|304160	product unknow... Bacillus subtilis]	54	50	489
502	2	1696	1133	gi|1205015	hypothetical protein (SP:P10120) [Haemophilus influenzae]	54	38	564
505	6	6262	5357	gi|1500558	2-hydroxyhepta-...,4-diene-1,7-dioate isomerase [Methanococcus jannaschii]	54	41	906
550	1	2736	1522	gi|40100	rodC (tagD) p...peptide (AA 1-746) [Bacillus subtilis] ir|S06049|S06049 rodC protei... - Bacillus subtilis p|P13485|TAGP_BACSU TECHOIC ACID BIOSYNTHES... (TEICH F.	54	35	1215
551	5	3305	4279	gi|950197	unknown [Cory... ...cterium glutamicum]	54	34	975
558	2	1356	958	gi|485090	No definiti... found [Caenorhabditis elegans]	54	32	399
580	1	91	936	gi|331906	fused envel... glycoprotein precursor [Friend spleen focus-forming virus]	54	45	846
603	3	554	757	gi|1323423	ORF YGR234w ... ...romyces cerevisiae]	54	36	204
617	1	25	249	gi|219959	ornithine tra... ...rbamylase [Homo sapiens]	54	40	225
622	3	1097	1480	gi|1303873	YqgZ [Bacill... ...ubtilis]	54	25	384
623	1	3	404	gi|1063250	low homology ... protein of Bacillus lichiniformis and bleomycin acetyltransf... ...t Streptomyces verticillus [Bacillus subtilis]	54	45	402
689	1	1547	1011	gi|552446	NADH dehydroge... ...ubunit 4 [Apis mellifera ligustica] pir|S52968|S52968 NADH dehydro... ...chain 4 - honeybee itochondrion [SGC4]	54	30	537
725	2	686	1441	gi|987096	sensory pro... ...ase [Streptomyces hygroscopicus]	54	26	756
956	1	1	249	pir|S30782|S307	integrin h... ...ast (Saccharomyces cerevisiae)	54	24	249
978	2	1137	859	gi|1301994	ORF YNL091w ... Saccharomyces cerevisiae]	54	33	279
1314	1	3	281	gi|1001108	hypothetic... ...ein [Synechocystis sp.]	54	33	279
2450	1	1	228	gi|1045057	ch-TOG [Hom... ...ens]	54	32	228
2934	1	1	387	gi|580870	ipa-37d go... ...ein [Bacillus subtilis]	54	36	387
2970	1	499	251	sp|P37348|YECE_	HYPOTHETIC... ...ASPS 5'REGION (FRAGMENT).	54	42	249

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene n. a.	% sim	% ident	length (nt)
3002	1	1	309	gi|34027	Tma protein (fr... ...occus lactis]	54	33	309
3561	1	9	464	gi|151259	HMG-CoA reduct... (EC 1.1.1.88) [Pseudomonas mevalonii] pir|A44756|A44756 hydroxymethyl...taryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	54	35	456
3572	1	72	401	gi|450688	...f gene product (Escherichia coli] pir|S38437|S38437 hsdM protein - Esc...richia coli pir|S09629|S09629 hypothetical protein A - Escherichia c... (SUB 40-520)	54	36	330
3829	1	798	400	gi|1322245	mevalonate py... ...diphosphate decarboxylase [Rattus norvegicus]	54	29	399
3909	1	1	273	gi|29865	CENP-E [Homo ...piens]	54	30	273
3921	1	3	209	pir|S24325|S243	glucan 1,4-beta-glucosidase (EC 3.2.1.74) - Pseudomonas fluorescens subsp. cellulosa	54	34	207
4438	1	566	285	gi|1196657	unknown protein [Mycoplasma pneumoniae]	54	30	282
4459	1	3	272	gi|1046081	hypothetical protein (GB:D26185_10) [Mycoplasma genitalium]	54	38	270
4564	1	3	221	gi|216267	ORF2 [Bacill... ...megaterium]	54	38	219
23	12	12538	10685	gi|474192	lucC gene pro... [Escherichia coli]	53	35	1854
23	14	14841	13579	gi|42029	ORF1 gene pr... ...Escherichia coli]	53	32	1263
24	3	4440	3940	gi|1369947	c2 gene produ... ...teriophage B1]	53	36	501
26	4	3818	4618	gi|1486247	unknown [Baci... ...btilis]	53	37	801
38	6	2856	3998	gi|405880	yeiI [Escher... ...oli]	53	40	1143
38	10	9380	7806	gi|1399954	thyroid sodi... ...e symporter NIS [Rattus norvegicus]	53	29	1575
56	10	12324	12100	pir|A54592|A545	110k actin fi... ...associated protein - chicken	53	32	225
57	6	5047	4583	pir|A00341|DEZP	alcohol dehy... ...s (EC 1.1.1.1) - fission yeast (Schizosaccharomyces pombe)	53	39	465
57	12	10515	8932	gi|480429	putative tra... ...ional regulator [Bacillus stearothermophilus]	53	30	1584
67	12	9496	10218	gi|1511555	quinolone ra... ...n rA protein (Methanococcus jannaschii]	53	31	723
69	3	3125	2382	gi|1687017	arabinogala... filtrate, Fr... ...AGP (Nicotiana alata, cell-suspension culture (61 aa]	53	30	744
79	1	3	1031	gi|1523802	glucanase [A... ...variabilis]	53	32	1029
80	1	673	338	gi|452428	ATPase 3 (Pl... ...falciparum]	53	36	336
88	4	1910	2524	gi|337034	ORF_o488 [Esc... ...coli]	53	25	615
88	5	2467	3282	gi|537034	ORF_o488 [Es... ...coli]	53	29	816

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
92	8	5870	5505	gi|399598	amphotropic murine retrovirus receptor [Rattus norvegicus]	53	33	366
94	5	4417	3239	gi|173038	tropomyosin TPM1 [Saccharomyces cerevisiae]	53	25	1179
99	5	4207	5433	sp|P28246|BCR_E	BICYCLOMYCIN RESISTANCE PROTEIN (SULFONAMIDE RESISTANCE PROTEIN).	53	30	1227
120	3	1639	2262	gi|576655	ORF1 [Vibrio anguillarum]	53	35	624
120	11	7257	8897	gi|1524397	glycine betaine transporter OpuD [Bacillus subtilis]	53	33	1641
127	6	6893	5685	gi|1256630	putative [Bacillus subtilis]	53	32	1209
147	2	255	557	gi|581648	epiB gene product [Staphylococcus epidermidis]	53	34	303
158	4	4705	4256	gi|151004	mucoidy regulatory protein AlgR [Pseudomonas aeruginosa] pir|A32802|A32802 regulatory protein algR - Pseudomonas aeruginosa sp|P26275|ALGR_PSEAE POSITIVE REGULATOR FOR ALGINATE BIOSYNTHESIS REGULATORY ROTEIN.	53	32	450
171	7	5717	5421	gi|1510669	hypothetical protein (GP:D64044_18) [Methanococcus jannaschii]	53	34	297
191	9	13087	11483	gi|298085	acetoacetate decarboxylase [Clostridium acetobutylicum] pir|B49346|B49346 butyrate--acetoacetate CoA-transferase (EC .8.3.9) small chain - Clostridium acetobutylicum sp|P33752|CTFA_CLOAB BUTYRATE-ACETOACETATE COA-TRANSFERASE SUBUNIT (EC 2.8.3.9) (COAT A)	53	31	1605
203	5	3763	4326	gi|143456	rpoE protein (ttg start codon) [Bacillus subtilis]	53	29	564
206	17	18204	18971	gi|304136	acetylglutamate kinase [Bacillus stearothermophilus] sp|Q07905|ARGB_BACST ACETYLGLUTAMATE KINASE (EC 2.7.2.8) (NAG INASE) (AGK) (N-ACETYL-L-GLUTAMATE PHOSPHOTRANSFERASE).	53	36	768
212	10	4021	4221	gi|9878	protein kinase [Plasmodium falciparum]	53	28	201
231	2	1580	1350	gi|537506	paramyosin [Dirofilaria immitis]	53	34	231
272	6	2719	3249	pir|A33141|A331	hypothetic [illegible] (gtfD 3' region) - Streptococcus mutans	53	34	531
308	3	927	2576	gi|606292	ORF_o696 [illegible]	53	33	1650
320	7	5645	5884	gi|160596	RNA polymerase [illegible] largest subunit [Plasmodium falciparum] sp|P27625 DNA-DIRECTED RNA POLYMERASE III LARGEST UBUNIT (EC 2.7.7.6).	53	33	240
327	1	218	901	gi|854601	unknown [Schizosaccharomyces pombe]	53	31	684
341	2	212	2500	gi|633732	ORF1 [Campylobacter jejuni]	53	31	2289
351	1	763	383	sp|P31675|YABH_	HYPOTHETICAL 42.7 KD PROTEIN IN TBPA-LEUD INTERGENIC REGION (ORF104).	53	32	381
433	7	5087	4731	gi|1001961	MHC class [illegible] [Staphylococcus aureus]	53	30	357
454	2	1240	980	pir|A60328|A603	40K cell [illegible] precursor (sr 5' region) - Streptococcus mutans (strain [illegible])	53	27	261

TABLE 2

S. aureus - Putative coding [illegible] of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	Match gene name	% sim	% ident	length (nt)
470	4	1123	1761	gi|516826	rat GCP360 [Rattus rattus]	53	30	639
483	1	432	217	gi|1480429	putative transcriptional regulator [Bacillus stearothermophilus]	53	33	216
544	1	516	1259	gi|46587	ORF 1 (AA 1 - [illegible] 1's 2nd base in codon) [Staphylococcus aureus] pir|S15765|S[illegible] hypothetical protein 1 (hlb 5' region) - aphylococcus aureus (fragment)	53	38	744
558	10	3957	3754	gi|15140	res gene [Bacteriophage P1]	53	32	204
603	2	339	620	gi|507738	Hmp [Vibrio parahaemolyticus]	53	26	282
693	1	1669	941	gi|153123	toxic shock [illegible] toxin-1 precursor [Staphylococcus aureus] pir|A24606|X[illegible]1 toxic shock syndrome toxin-1 precursor - taphylococcus aureus	53	38	729
766	1	2	673	gi|687600	orfA2; orfA2 [illegible] an operon with orfA1 [Listeria monocytogenes]	53	43	672
781	1	667	335	gi|1204551	pilin biogen[illegible] protein [Haemophilus influenzae]	53	26	333
801	1	3	545	gi|1279400	SapA protein [Escherichia coli]	53	25	543
803	1	2	910	gi|695278	lipase-like [illegible] Alcaligenes eutrophus]	53	30	909
872	1	1177	590	gi|298032	EP [Streptoc[illegible]]	53	30	588
910	1	2	184	gi|1044936	unknown [Sch[illegible]ar myces pombe]	53	29	183
943	1	794	399	gi|390508	similar to u[illegible] [illegible] of near 47 minutes [Escherichia coli] sp|P31436|Y[illegible]001 HYPOTHETICAL 43.5 KD PROTEIN IN SELC-NLPA NTERGENIC REGION.	53	30	396
988	1	1004	504	gi|142441	ORF 3; putat[illegible] [illegible]illus subtilis]	53	28	501
1064	1	3	434	gi|305080	myosin heavy [illegible] [Entamoeba histolytica]	53	26	432
1366	1	3	452	gi|308852	transmembran[illegible] [illegible] [Lactococcus lactis]	53	33	450
1758	1	792	397	gi|1001774	hypothetical [illegible] [Synechocystis sp.]	53	30	396
1897	1	1	447	gi|1303949	YqiX [Bacil[illegible]]	53	27	447
2381	1	798	400	gi|1146243	22.4% identi[illegible] [illegible] Escherichia coli DNA-damage inducible protein ...; putative [E[illegible]is]	53	37	399
3537	1	1	327	gi|450688	hsdM gene of [illegible] product [Escherichia coli] pir|S38437|S38437 hsdM protein - E[illegible] coli pir|S09629|S09629 hypothetical protein A - Escherichi[illegible] (SUB 16 520)	53	35	327
3747	2	137	397	gi|1477486	transposase [illegible] [illegible] [illegible]epacia]	53	53	261
11	5	3049	3441	gi|868224	No definiti[illegible] [illegible] [illegible] Caenorhabditis elegans]	52	33	393

TABLE 2

S. aureus - Putative coding [illegible] of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene [illegible]	% sim	% ident	length (nt)
15	5	2205	2369	gi|215966	G41 protein [gtg start codon) [Bacteriophage T4]	52	34	165
19	3	2429	3808	gi|1205379	UDP-murnac-[illegible] synthetase [Haemophilus influenzae]	52	31	1380
24	1	6920	3462	gi|579124	predicted 86.4kd protein; 52Kd observed [Mycobacteriophage 15] pir|S30971|S30971 gene 26 protein - Mycobacterium phage L5 sp|Q05233|VG26_BPML5 MINOR TAIL PROTEIN GP26. (SUB 2-837)	52	32	3459
37	5	3015	3935	gi|1500543	P115 protein [Methanococcus jannaschii]	52	25	921
38	13	8795	9703	gi|46851	glucose kinase [Streptomyces coelicolor]	52	29	909
44	16	10617	11066	gi|42012	moaE gene product [Escherichia coli]	52	36	450
46	1	3	521	gi|1040957	NADH dehydro[illegible]nase [illegible]unit 6 [Anopheles trinkae]	52	25	519
51	10	5531	6280	gi|388269	traC [Plasm.[illegible] pAM]	52	32	750
56	5	3968	2826	gi|181949	endothelial [illegible] protein (edg-1) [Homo sapiens] pir|A35300|[illegible] in-coupled receptor edg-1 - human sp|P21453|[illegible] ABLE G PROTEIN-COUPLED RECEPTOR EDG-1.	52	23	1143
57	5	4850	4173	gi|304153	sorbitol del[illegible]rogena[illegible] [Bacillus subtilis]	52	27	678
62	5	3364	2870	gi|1072399	phaE gene p[illegible] [illegible]rium meliloti]	52	25	495
62	6	4445	3651	gi|46485	NADH dehydr[illegible]genas[illegible] [Synechococcus PCC7942]	52	27	795
67	14	11355	12962	gi|1511365	glutamate s[illegible] (N-D[illegible]), subunit alpha [Methanococcus jannaschii]	52	30	1608
67	21	16935	18158	gi|1204393	hypothetical protein (SP:P31122) [Haemophilus influenzae]	52	25	1224
70	4	2185	1997	gi|7227	cytoplasmic [illegible] heavy chain [Dictyostelium discoideum] r|A44357|A44357 dynein hea[illegible] chain, cytosolic - slime mold ctyostelium discoideum)	52	36	189
96	10	10005	10664	gi|1408485	B65G gene p[illegible] [illegible]illus subtilis]	52	26	660
103	5	3986	3351	gi|1009368	Respiratory [illegible] [illegible]ductase [Bacillus subtilis]	52	42	636
109	3	4102	3350	gi|699274	lmbE gene [illegible] [illegible]Mycobacterium leprae]	52	39	753
109	19	15732	17300	gi|1526981	amino acid [illegible] like protein [Salmonella typhimurium]	52	30	1569
121	3	1412	981	gi|732931	unknown [Sa[illegible] ar[illegible] evisiae]	52	32	432
125	3	865	1680	gi|1296975	puT gene pr[illegible] [illegible]monas gingivalis]	52	38	816
130	2	659	1807	gi|1256634	25.8% ident[illegible] with the Synenococcus sp. HpeV protein; putative [Bacillus [illegible]]	52	36	1149
149	1	1164	583	gi|1225943	PBSX termin[illegible] [illegible]s subtilis]	52	33	582
149	14	4687	4415	gi|1510368	H. jannasch[illegible] [illegible]ding region MJ0272 [Methanococcus jannaschii]	52	35	273

TABLE 2

S. aureus - Putative coding [illegible] of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
167	1	216	1001	gi|146025	cell division protein [Escherichia coli]	52	43	786
188	1	120	1256	gi|474915	orf 337; translated similarity to SW: BCR_ECOLI bicyclomycin esistance protein of Escherichia coli [Coxiella burnetii] pir|S44207|S44207 hypothetical protein 337 - Coxiella burnetii (SUB -338)	52	26	1137
195	9	9161	8760	gi|3028	mitochondrial outer membrane 72K protein [Neurospora crassa] r|A36682|A36682 72K mitochondrial outer membrane protein - Neurospora crassa	52	25	402
200	3	2065	2607	gi|142439	ATP-dependent nuclease [Bacillus subtilis]	52	35	543
203	4	2776	3684	gi|1303698	BltD [Bacillus subtilis]	52	25	909
227	8	5250	5651	gi|305080	myosin heavy chain [Entamoeba histolytica]	52	24	402
242	1	21	1424	gi|1060877	EmrY [Escherichia coli]	52	32	1404
249	5	4526	4753	pir|C37222|C372	cytochrome P4[illegible] hepatic - dog (fragment)	52	23	228
255	1	2107	1055	gi|143290	penicillin-binding protein [Bacillus subtilis]	52	28	1053
276	7	3963	3664	gi|1001610	hypothetical protein [Synechocystis sp.]	52	30	300
276	8	4456	4055	gi|416235	orf L3 [Mycoplasma [illegible]ium]	52	26	402
289	2	1856	1449	gi|150900	GTP phosphohydrolase [Hordeum vulgaris]	52	34	408
325	1	1	279	gi|1204874	polypeptide deformylase (formylmethionine deformylase) [Haemophilus influenzae]	52	33	279
340	1	2017	1010	gi|1215695	peptide transport system protein SapF homolog; SapF homolog [Mycoplasma pneumoniae]	52	33	1008
375	3	340	1878	gi|467446	similar to SpoVB [Bacillus subtilis]	52	28	1539
424	4	4104	3262	gi|1478239	unknown [Mycobacterium tuberculosis]	52	34	843
430	1	3	575	pir|A42606|A426	orfA 5' to ori[illegible] Saccharopolyspora erythraea (fragment)	52	28	573
444	4	4728	3712	gi|1408494	homologous to [illegible]n acylase [Bacillus subtilis]	52	31	1017
465	1	1802	903	gi|143331	alkaline phos[illegible] regulatory protein [Bacillus subtilis] pir|A27650|A[illegible] regulatory protein phoR - Bacillus subtilis sp|P23545|PH[illegible]ALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN HOR (EC 2.7.3.-).	52	36	900
469	5	4705	4169	gi|755152	highly hydrop[illegible]t membrane protein [Bacillus subtilis] sp|P42953|TA[illegible]HOIC ACID TRANSLOCATION PERMEASE PROTEIN AGG.	52	32	537
495	1	1262	633	gi|1204607	transcription [illegible]aemophilus influenzae]	52	25	630
505	7	6004	5762	gi|142440	ATP-dependent [illegible]acillus subtilis]	52	28	243

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	matdh gene	% sim	% ident	length (nt)
517	2	1162	1614	gi|166162	Bacteriophage phi-11 int gene activator (Staphylococcus acteriophage phi 11)	52	35	453
543	2	444	1295	gi|1215693	putative ort. Orf_ [Mycoplasma pneumoniae]	52	25	852
586	1	1	336	gi|581648	epiB gene pr [Staphylococcus epidermidis]	52	36	336
773	1	848	426	gi|1279769	FdhC [Methan [Thermoformicicum]	52	30	423
1120	2	100	330	gi|142439	ATP-depende se [Bacillus subtilis]	52	35	231
1614	1	691	347	gi|289262	comE ORF3 [B ubtilis]	52	28	345
2495	1	1	324	gi|216151	DNA polymerase (g ttg start codon) [Bacteriophage SPO2] gi|579197 SPO2 DNA pol [aa 1-648] [Bacteriophage SPO2] pir|A21498|DJBPS2 DNA-directed DN se (EC 2.7.7.7) - phage PO2	52	34	324
2931	1	566	285	gi|1256136	YbbG [Bacil]	52	30	282
2943	1	577	320	gi|41713	hisA ORF (Escherichia coli]	52	35	258
2993	1	588	295	gi|298032	EF (Strepto s]	52	34	294
3667	1	612	307	gi|849025	hypothetica rotein [Bacillus subtilis]	52	36	306
3944	1	478	260	gi|1218040	BAA [Bacil mis]	52	36	219
3954	2	613	347	gi|854064	U87 [Human]	52	50	267
3986	1	90	401	gi|1205919	Na+ and Cl- ma-aminobutryic acid transporter [Haemophilus influenza	52	33	312
4002	1	3	389	gi|40003	oxoglutarat (NADP+) [Bacillus subtilis] p|P23129|ODO1_BACSU 2-OXOGLUTA GENASE E1 COMPONENT (EC 2.4.2) (ALPHA- KETOGLUTARATE DEHYDROGEN	52	42	387
4020	1	1	249	gi|159388	ornithine d [Leishmania donovani]	52	47	249
4098	1	438	220	gi|409795	No definiti [Escherichia coli]	52	32	219
4248	1	3	212	gi|965077	Adr6p [Sac visiae]	52	40	210
7	1	3	575	gi|895747	putative c ator [Bacillus subtilis]	51	28	573
21	4	2479	3276	gi|1510962	indole-3-gl hate synthase [Methanococcus jannaschii]	51	32	798
22	9	5301	5966	gi|1303933	YqiN [Baci]	51	25	666
43	3	1516	1283	gi|1519460	Srp1 [Schiz pombe]	51	31	234
44	17	11042	11305	gi|42011	moaD gene richia coli]	51	35	264
51	11	6453	6731	gi|495471	vacuolatin ter pylori]	51	37	279

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
52	4	2537	2995	gi|1256652	25% identity to the E.coli regulatory protein MprA; putative [Bacillus subtilis]	51	32	459
57	10	7331	6843	gi|508173	EIIA domain of PTS-dependent Gat transport and phosphorylation Escherichia coli]	51	32	489
59	1	29	1111	gi|299163	alanine dehydrogenase (Bacillus subtilis)	51	33	1083
67	20	15791	16576	gi|1510977	M. jannaschii predicted coding region MJ0938 [Methanococcus jannaschii]	51	24	786
69	2	1559	1218	gi|467359	unknown [Bacillus subtilis]	51	34	342
71	1	3	1196	gi|298032	EF [Streptococcus suis]	51	32	1194
78	2	349	176	gi|1161242	proliferating cell nuclear antigen [Styela clava]	51	28	174
99	4	3357	4040	gi|642795	TFIID subunit TAFII95 [Homo sapiens]	51	25	684
109	1	2852	1428	gi|580920	rodD (gtaA) protein (AA 1-673) [Bacillus subtilis] pir|S06048|S06048 probable rodD protein - Bacillus subtilis sp|P13484|TAGE_BACSU PROBABLE POLY(GLYCEROPHOSPHATE) ALPHA-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHOIC ACID BIOSYNTHESIS PROTEIN E).	51	27	1425
109	9	6007	6693	gi|1204815	hypothetical protein (SP:P2.662) [Haemophilus influenzae]	51	23	687
112	3	1066	2352	pir|S05330|S053	maltose-binding protein precursor - Enterobacter aerogenes	51	42	1287
112	13	14432	12855	gi|405857	yehU [Escherichia coli]	51	29	1578
114	9	9725	8967	gi|435098	orf1 [Mycoplasma hyorhinis]	51	30	759
115	1	1	912	gi|1431110	ORF YDL085w [Saccharomyces cerevisiae]	51	28	912
127	10	9647	10477	gi|1204314	H. influenzae predicted coding region HI0056 [Haemophilus influenzae]	51	37	831
152	9	6814	7356	gi|431929	HunI regulatory protein [Mycoplasma sp.]	51	38	543
154	2	575	1153	gi|1237044	unknown [Mycobacterium tuberculosis]	51	36	579
154	7	6587	5634	gi|409286	bmrU [Bacillus subtilis]	51	27	954
171	8	6943	6236	gi|1205484	hypothetical protein (SP:P43718) [Haemophilus influenzae]	51	32	708
184	1	1	291	gi|466886	B1496_C3_206 [Mycobacterium leprae]	51	33	291
212	5	1501	2139	pir|A45605|A456	mature-parasite-infected erythrocyte surface antigen MESA - Plasmodium falciparum	51	23	639
228	2	707	1378	gi|8204	nuclear protein [Drosophila melanogaster]	51	27	672
236	8	8137	7481	gi|49272	Asparaginase [Bacillus licheniformis]	51	31	657
243	4	4637	3546	gi|1511102	mevalonate kinase [Methanococcus jannaschii]	51	29	1092

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match Accession	match gene name	% sim	% ident	length (nt)
257	4	3540	3373	gi|204579	H. influenz… coding region HI0326 [Haemophilus influenzae]	51	22	168
258	3	2397	1609	gi|160299	glutamic ac… protein [Plasmodium falciparum] pir|A54514|A54514 glutamic ac… protein precursor - Plasmodium alciparum	51	34	789
265	5	2419	3591	gi|580841	F1 [Bacillus … is]	51	32	1173
298	2	518	748	gi|1336162	SCPB [Strep… agalactiae]	51	34	231
316	9	5817	7049	gi|413953	ipa-29d gen… product [Bacillus subtilis]	51	39	1233
332	2	3775	2057	gi|209012	mutS [Therm… …icus thermophilus]	51	26	1719
364	4	3816	4991	gi|528991	unknown [Ba… subtilis]	51	32	1176
440	2	448	684	gi|2819	transferase … (AA 1 - 687) [Kluyveromyces lactis] r|S01407|XUVKG UDPglucos… …se (EC 5.1.3.2) - yeast uyveromyces marxianus var. lactis)	51	32	237
495	2	1353	1177	gi|297861	protease G … chrysanthemi]	51	41	177
495	3	2287	1718	gi|1513317	serine rich … entamoeba histolytica]	51	25	570
506	1	840	421	gi|455320	cII protei… …P4]	51	33	420
600	1	1474	983	gi|587532	orf, len: 2… …[Saccharomyces cerevisiae] pir|S48818|S48818 hypothetic… …east (Saccharomyces erevisiae)	51	30	492
607	3	479	934	gi|1511524	hypothetic… …in (SI:P37002) [Methanococcus jannaschii]	51	40	456
686	2	127	600	gi|493017	endocardit… …ic antigen [Enterococcus faecalis]	51	30	474
726	1	33	230	gi|1353851	unknown [Pr… …coccus …arinus]	51	45	198
861	1	176	652	gi|410145	dehydroqui… …ydrase [Bacillus subtilis]	51	34	477
869	1	782	393	gi|40100	rodC (tag) … …de (AA 1-746) [Bacillus subtilis] ir|S06049|S06049 rodC prot… …illus subtilis p|P13485|TAGF_BACSU TECHOIC ACID BIOSYNTHE… …F.	51	23	390
1003	1	642	322	gi|1279707	hypothetic… …ycerate mutase (Saccharomyces cerevisiae)	51	39	321
1046	2	866	624	gi|510257	glycosylt… …Escherichia coli]	51	29	243
1467	1	702	352	gi|1511175	M. jannasc… …ed coding region MJ1177 [Methanococcus jannaschii]	51	32	351
2558	1	457	230	sp|P10582|DPOH_	DNA POLYME… …) (S-1 DNA ORF 3).	51	26	228
3003	1	779	399	gi|809543	CbrC prote… …chrysanthemi]	51	27	381
3604	1	1	399	pir|JC4210|JC42	3-hydroxya… …acyl…genase (EC 1.1.1.35) - mouse	51	37	399
3732	1	2	316	gi|145906	acyl-CoA d… …e [Escherichia coli]	51	33	315

TABLE 2

S. aureus - Putative codin[illegible] [illegible]ions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene [illegible]	% sim	% ident	length (nt)
3791	1	2	274	gi|1061351	semaphorin [illegible] [illegible]molog [Homo sapiens]	51	37	273
3995	1	46	336	gi|216346	surfactin sy[illegible] [illegible]acillus subtilis]	51	38	291
4193	1	612	307	gi|42749	ribosomal p[illegible] (AA 1-179) [Escherichia coli] ir|S04776|XXECPL peptide N-[illegible] [illegible] rimL (EC 2.3.1.-) - cherichia coli	51	25	306
4539	1	367	185	gi|1408494	homologous [illegible] [illegible]enicillin [illegible]cylase [Bacillus subtilis]	51	40	183
4562	1	442	239	gi|1458280	coded for by [illegible] elegans cDNA cm01e7; Similar to hydroxymethylglutaryl-CoA synthase [C[illegible]norhabditis elegans]	51	35	204
1	4	3576	4859	gi|559160	GRAIL score: [illegible]ll; cap site and late promoter motifs present pstream; putative [A[illegible] [illegible] californica nuclear polyhedrosis irus]	50	44	1284
11	7	4044	5165	gi|1146207	putative [B[illegible] [illegible]us subtilis]	50	35	1122
11	13	10509	9496	gi|1208451	hypothetica[illegible] [illegible] in [Synechocystis sp.]	50	39	1014
19	1	2034	1018	gi|413966	ipa-42d gen[illegible] [illegible] [Bacillus subtilis]	50	29	1017
20	11	8586	8407	gi|1323159	ORF YGR103w [illegible]myces cerevisiae]	50	28	180
24	5	5408	4824	gi|496280	structural [illegible] [illegible]ophage Tuc2009]	50	29	585
34	4	1926	2759	gi|1303966	YqjO [Bacil[illegible]	50	36	834
38	30	22865	23440	gi|1072179	Similar to [illegible] 4-reductase (maize, petunia, tomato) [Caenorha[illegible]	50	32	576
47	2	1705	2976	gi|153015	FemA protei[illegible] [illegible]cus aureus]	50	29	1272
56	13	15290	15841	gi|606096	ORF_f167; [illegible] [illegible] of o100 by 14 bases; start overlaps f174, ther starts pos[illegible] [illegible]herichia coli]	50	30	552
57	1	2135	1077	gi|640922	xylitol deh[illegible] [unidentified hemiascomycete]	50	29	1059
58	2	628	1761	gi|143725	putative [B[illegible] [illegible]btilis]	50	29	1134
88	6	4393	3884	gi|1072179	Similar to [illegible] [illegible]vonol-4-reductase (maize, petunia, tomato) [Caenorha[illegible] [illegible]ns]	50	32	510
89	5	3700	3356	gi|1276658	ORF174 gen[illegible] [illegible] purpurea]	50	25	345
141	1	3	239	gi|476024	carbamoyl p[illegible] [illegible]ase II [Plasmodium falciparum]	50	33	237
151	1	186	626	gi|1403441	unknown [M[illegible] [illegible]rculosis]	50	35	441
166	7	11065	9623	gi|895747	putative c[illegible] [illegible]or [Bacillus subtilis]	50	32	1443
201	6	5284	5096	gi|160229	circumspor[illegible] [illegible]lasmodium reichenowi]	50	42	189
206	22	30784	29555	gi|1052754	LarP integ[illegible] [illegible]tein [Lactococcus lactis]	50	24	1230

TABLE 2

S. aureus - Putative coding [illegible] proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene na[illegible]	% sim	% ident	length (nt)
211	4	1523	1927	gi|410131	ORFX7 [Bacillu[illegible]	50	29	405
214	4	2411	3295	sp|P37348|YECE_	HYPOTHETICAL H[illegible] 5'REGION (FRAGMENT).	50	37	885
228	7	5068	4406	gi|313580	envelope prot[illegible] [illegible]odeficiency virus type 1] pir|S35835|S35835 envelope pro[illegible] [illegible]unodeficiency virus ype 1 (fragment) (SUB 1-77)	50	35	663
272	2	3048	1723	gi|1408485	B65G gene pro[illegible] lo subtilis]	50	22	1326
273	2	1616	984	gi|JH4186	phosphoglycer[illegible] haromyces cerevisiae]	50	28	633
328	2	2507	1605	gi|148896	lipoprotein [illegible] luenzae]	50	26	903
332	4	5469	3802	gi|1526547	DNA polymeras[illegible] [illegible] rmus aquaticus]	50	27	1668
342	5	3473	3931	gi|456562	G-box binding [illegible] stelium discoideum]	50	35	459
352	1	1478	741	gi|288301	ORF2 gene pro[illegible] megaterium]	50	29	738
408	7	5299	5523	gi|11665	ORF2136 [Marc[illegible] monas]	50	27	225
420	3	650	1825	gi|757842	UDP-sugar hyd[illegible] lia coli]	50	30	1176
464	1	1	591	gi|487282	Na+ -ATPase s[illegible] coccus hirae]	50	29	591
472	2	1418	864	gi|551875	BglR [Lactocu[illegible]	50	23	555
520	1	23	541	gi|567036	CapE [Staphy[illegible]	50	27	519
529	1	6	410	gi|1256652	25% identity [illegible] regulatory protein MprA; putative [Bacillus subtilis]	50	34	405
534	5	7726	6059	gi|295671	selected as [illegible] of a mutant of the subunit AC40 of DNA dependant RNA [illegible] d III [Saccharomyces cerevisiae]	50	18	1668
647	1	2990	1497	gi|405568	TraI protein [illegible] similarity with a family of opoisomerases [Plasmid pSA[illegible]	50	31	1494
664	3	1133	711	gi|410007	leukocidin F [illegible] ylococcus aureus, MRSA No. 4, Peptide, 23 aa]	50	32	423
678	1	1	627	gi|238032	EF (Streptoc[illegible]	50	29	627
755	3	947	1171	gi|150572	cytochrome c1 [illegible] .10.2.2) [Paracoccus denitrificans] gi|45465 cytochrome c[illegible] Paracoccus denitrificans] pir|C29413|C29413 ubiquinol--c[illegible] ctase (EC 1.10.2.2) ytochrome c1 precursor - Paracoccus d[illegible] P13627|CY1	50	37	225
827	1	1363	683	gi|142020	heterocyst di[illegible] otein [Anabaena sp.]	50	21	681
892	1	3	752	gi|1408485	B65G gene pr[illegible] btilis]	50	27	750
910	2	438	887	gi|1104727	tyrosine-spec[illegible] rotein [Haemophilus influenzae]	50	25	450

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding [illegible] novel proteins similar to known proteins

Contig ID	Diff ID	Start (nt)	Stop (nt)	match accession	match gene [illegible]	% sim	% ident	length (nt)
933	1	524	760	gi|1205451	cell division [illegible] of Haemophilus influenzae]	50	32	237
973	1	424	236	gi|886947	orf3 gene pr[illegible] Saccharomyces cerevisiae]	50	40	189
1009	1	653	429	gi|153727	M protein [S. pyogenes]	50	28	225
1027	1	511	257	gi|413934	ipa-10r gene [Bacillus subtilis]	50	25	255
1153	2	556	326	gi|773676	nccA [Alcaligenes sp.]	50	36	231
1222	1	798	400	gi|1408485	B65G gene pr[illegible] subtilis]	50	21	399
1350	1	692	399	gi|289272	ferrichrome[illegible] [Bacillus subtilis]	50	32	294
2945	1	366	184	gi|171704	hexaprenyl[illegible] synthetase (COQ1) [Saccharomyces cerevisiae]	50	34	183
2968	2	1604	804	gi|397526	clumping fa[illegible] Staphylococcus aureus]	50	33	801
2998	2	657	394	gi|495696	F54E7.3 gene[illegible] Caenorhabditis elegans]	50	40	264
3046	2	506	306	pir|S13819|S138	acyl carrier[illegible] protein variabilis (fragment)	50	32	201
3063	1	547	275	gi|174190	iucA gene p[illegible] Escherichia coli]	50	29	273
3174	1	3	146	gi|151900	alcohol deh[illegible] Zymobacter sphaeroides]	50	31	144
3792	1	625	314	gi|1001423	hypothetica[illegible] Synechocystis sp.]	50	35	312
3800	1	2	262	gi|144733	NAD-depende[illegible] acetobutyl[illegible] butyryl coenzyme A dehydrogenase Clostridium	50	28	261
3946	1	373	188	gi|576765	cytochrome[illegible] [Sulfolobus sula]	50	38	186
3984	1	578	291	sp|P37348|YECE_	HYPOTHETICA[illegible] PE 5' REGION (FRAGMENT).	50	37	288
37	10	8250	7885	gi|1204367	hypothetica[illegible] (U00003_279) [Haemophilus influenzae]	49	30	366
46	16	13802	14848	gi|466860	acd; B1308_[illegible] [Mycobacterium leprae]	49	24	1047
59	5	2267	3601	gi|606304	ORF_o462 [E[illegible]	49	27	1335
112	18	17884	18615	gi|559502	ND4 protein[illegible] [Caenorhabditis elegans]	49	25	732
138	9	6973	7902	gi|303953	esterase [A[illegible] [Psychrobacter immobilis]	49	29	930
217	6	4401	5138	gi|496254	fibronecti[illegible] binding protein [Streptococcus pyogenes]	49	31	738
220	12	11803	12657	gi|397526	clumping f[illegible] [Staphylococcus aureus]	49	31	855
228	4	1842	2492	pir|S23692|S236	hypothetic[illegible] [Plasmodium falciparum]	49	24	651
268	1	5016	2614	gi|143047	ORF8 [Baci[illegible]	49	26	2403

TABLE 2

S. aureus - Putative coding [illegible] proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene [illegible]		% sim	% ident	length (nt)
271	2	1164	1373	gi|1001257	hypothetical	[illegible]ocystis sp.]	49	38	210
300	3	4340	3180	gi|1510796	hypothetical	[illegible]1006_2] (Methanococcus jannaschii)	49	26	1161
381	1	2281	1142	gi|396301	matches PS00 [illegible] (Escherichi	[illegible]egulatory proteins, araC family signature	49	29	1140
466	1	3	947	gi|1303863	YqgP (Bacill	[illegible]	49	26	945
666	1	379	191	gi|633112	ORF1 (Strep	[illegible]	49	29	189
670	2	403	1014	gi|1122758	unknown (Bac	[illegible]	49	32	612
709	1	1433	795	gi|114830	xpaC (Bacill	[illegible]	49	29	639
811	1	943	473	gi|401786	phosphomann	[illegible]ida pirum]	49	29	471
1052	1	422	213	gi|1303799	YqeN (Bacill	[illegible]	49	21	210
1800	1	342	172	gi|216300	peptidoglyc HURG PROTE PENTAPEPTI	[illegible]ype (Bacillus subtilis) sp|P37585|MURG_BACSU GLUCOSAMINE--N-ACETYLMURAMYL- [illegible]D-UNDECAPRENOL N-ACETYLGLUCOSAMINE RANSFERASE).	49	28	171
2430	1	2	376	sp|P27434|YFCA_	HYPOTHETICA	[illegible]N IN NDK-GCPE INTERGENIC REGION.	49	26	375
3096	1	542	273	gi|516360	surfactin s	[illegible]cus subtilis]	49	25	270
32	4	3771	3100	gi|1217963	hepatocyte	[illegible] gamma (IRF4gamma) (Homo sapiens]	48	36	672
38	1	1	609	gi|1205790	H. influenz	[illegible]ng region HI1555 (Haemophilus influenzae)	48	28	609
45	6	5021	6427	gi|1524267	unknown (My	[illegible]culosis]	48	20	1407
59	14	16346	31096	gi|1197336	Lmp3 protei	[illegible]anis]	48	28	14751
61	1	3	608	gi|1511555	quinolone r	[illegible]rotein protein (Methanococcus jannaschii]	48	30	606
61	3	3311	3646	gi|1303893	YqhL (Bacil	[illegible]	48	29	336
114	1	98	415	gi|671708	su(s) homol (su(s)) pr	[illegible]rosophila melanogaster suppressor of able t Accession Number P22293 Drosophila virilis)	48	25	318
121	1	1131	610	gi|1314584	unknown (Sy	[illegible]	48	29	522
136	1	2014	1280	gi|1205968	H. influenz	[illegible]ing region HI1738 (Haemophilus influenzae)	48	23	735
171	10	8220	9557	gi|1208454	hypothetica	[illegible]nocystis sp.]	48	34	1338
175	1	3625	1814	gi|396400	similar to sp|P32703| REGION 105	[illegible] exchangers [Escherichia coli] VERTICAL 60.5 KD PROTEIN IN SOXR-ACS NTERGENIC	48	29	1812
194	1	2	385	gi|1510493	H. jannasch	[illegible]ing region MJ0419 (Methanococcus jannaschii]	48	25	384

TABLE 2

S. aureus - Putative codi... ...vel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene	% sim	% ident	length (nt)
197	1	901	452	gi|1045714	spermidine/ ...n port ATP-binding protein [Mycoplasma genitalium]	48	25	450
203	1	1	396	gi|940288	protein [oc ... sativum]	...ucleoli of pea nuclei; ORF; putative Pisum	48	29	396
204	1	1363	698	gi|529202	No definit...	...aenorhabditis elegans]	48	25	666
206	20	34815	27760	gi|511490	gramicidin	...Bacillus brevis]	48	27	7056
212	1	2	166	gi|295899	nucleolin [...		48	34	165
220	10	12652	11426	gi|44073	SecY protein ...	...lactis]	48	23	1227
243	6	6450	5491	gi|1184118	mevalonate ...	...bacterium thermoautotrophicum]	48	30	960
264	4	5434	3308	gi|1015903	ORF YJR151c ...	...cerevisiae]	48	26	2127
441	1	1532	768	gi|142863	replicatio... ... replicati...	...ein [Bacillus subtilis] pir|B26580|B26580 ...rotein - Bacillus ubtilis	48	23	765
444	5	3898	5298	gi|145836	putative [... ...]		48	24	1401
484	2	388	1110	gi|146551	transmembr... ...	...D) [Escherichia coli]	48	18	723
542	3	1425	2000	pir|S28969|S289	N-carbamoyl ...	...ydrolase (EC 3.5.1.59) - Arthrobacter sp.	48	27	576
566	1	3	1019	gi|153490	tetracenom...	...e and export protein [Streptomyces laucascens]	48	24	1017
611	1	2	730	gi|1103507	unknown [S...	...s pombe]	48	38	729
624	1	1255	665	gi|144859	ORF B [Clos...	...ens]	48	26	591
846	1	1014	508	gi|537506	paramyosin ...	...ilis]	48	27	507
1020	1	66	950	gi|1499876	magnesium ...	...port protein [Methanococcus jannaschii]	48	30	885
1227	1	1	174	gi|493730	lipoxygenase ...	...d	48	35	174
1266	1	1	405	gi|882452	ORF_f211; ... ORF5 (AA ...	...gA; orf5 of X14436 [Escherichia coli] gi|41425 ...ia coli] [SUB 15-211]	48	24	405
2071	1	707	381	gi|1408486	HS74A gen...	...s subtilis]	48	25	327
2398	1	463	233	gi|1500401	reverse gy...	...us jannaschii]	48	40	231
2425	1	476	246	pir|H48563|H485	GI protein ...	...(strain HP444) (fragment)	48	40	231
2432	1	446	225	gi|1353703	Trio [Hom...		48	33	222
2453	1	794	399	gi|142850	division i...	...[Bacillus subtilis]	48	29	396
2998	1	469	236	gi|577569	PepV [Lact...	...ckii]	48	31	234

TABLE 2

S. aureus - Putative coding [illegible] proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene n[illegible]		% sim	% ident	length (nt)
3042	1	14	280	gi|945219	mucin [Homo s[illegible]		48	35	267
3686	1	1	405	gi|145836	putative [Esc[illegible]	[illegible]col]	48	25	405
4027	2	492	301	pir.S51177|S511	trans-activat[illegible]	[illegible]n - [illegible]quine infectious anemia virus	48	32	192
4	2	3641	2232	gi|1303989	YqkI [Bacillu[illegible]	[illegible]is]	47	24	1410
24	2	599	1084	gi|540083	PC4-1 gene pr[illegible]	[illegible], [illegible]ia hygida]	47	28	486
36	10	7524	6925	gi|1209223	esterase [Ac[illegible]	[illegible]ftii]	47	26	600
43	2	196	1884	gi|1403455	unknown [Myco[illegible]	[illegible], [illegible]culosis]	47	27	1689
44	22	16118	15108	gi|1511555	quinolone re[illegible]	[illegible]otein protein [Methanococcus jannaschii]	47	31	1011
69	7	7141	6710	gi|438466	Possible ope[illegible] putative [B[illegible]	[illegible]ydrophilic, no homologue in the atabase; [illegible]s]	47	29	432
81	4	5022	4279	gi|466882	ppsl; B1496_C[illegible]	[illegible]terium leprae]	47	24	744
120	12	9135	8863	gi|927340	D9509.27p; CA[illegible]	Sa[illegible]romyces cerevisiae]	47	38	273
142	1	2022	1174	gi|486143	ORF YKL094w [illegible]	myc[illegible] cerevisiae]	47	32	849
168	1	2178	1093	gi|1177254	hypothetical [illegible]	[illegible]Bacillus subtilis]	47	29	1086
263	1	1884	943	gi|142822	D-alanine ra[illegible]	[illegible]illus subtilis]	47	34	942
279	1	1109	561	gi|516608	2 predicted [illegible] et. al. unp[illegible] updated Ru[illegible] HYPOTHETICA[illegible]	[illegible]s, homology with B. subtilis man Orf3 Rowland [illegible]ion number M74183), approximately 1 minutes on [illegible]e [Escherichia coli] sp|P37355|YFBB_ECOLI [illegible]N IN MEND-MENB	47	31	549
345	2	2620	1676	gi|1204835	hippuricase [illegible]	[illegible]luenzae]	47	28	945
389	2	152	400	gi|456562	G-box bindin[illegible]	[illegible], [illegible]stelium discoideum]	47	32	249
391	1	1	831	gi|1420856	myo-inositol [illegible]	[illegible]izosaccharomyces pombe]	47	19	831
404	3	2072	2773	gi|1255425	C33G8.2 gene [illegible]	[illegible]rnabditis elegans]	47	17	702
529	5	2145	3107	gi|1303973	YqjV [Bacill[illegible]	[illegible]	47	29	963
565	2	2321	1257	gi|142824	processing p[illegible]	[illegible]s subtilis]	47	28	1065
654	1	962	483	gi|243353	ORF 5' of EC[illegible] aa]	[illegible]v[illegible]s saimiri HVS, host-squirrel monkey, eptide, 407	47	23	480
692	1	115	633	gi|150756	40 kDa prote[illegible]	[illegible]d[illegible]]	47	25	519
765	1	1634	819	gi|1256621	26.7% of ide[illegible] protein 6; [illegible]	[illegible]to a Thermophilic bacterium hypothetical [illegible]us subtilis]	47	28	816

TABLE 2

S. aureus - Putative codi[illegible] of [illegible]vel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
825	2	211	1023	gi|397526	clumping [illegible] ccus aureus]	47	32	813
914	1	1	615	gi|558073	polymorphic [illegible] dium falciparum]	47	29	615
1076	1	1	753	gi|1147557	Aspartate [illegible] [Bacillus circulans]	47	33	753
1351	1	793	398	gi|755153	ATP-binding [illegible] us subtilis]	47	20	396
4192	1	3	293	gi|145836	putative [Es[illegible]	47	24	291
5	6	4708	4361	gi|305080	myosin heavy [illegible] ba histolytica]	46	30	348
11	4	2777	3058	gi|603639	Yel040p [Sa[illegible] visiae]	46	28	282
46	11	10518	10300	gi|1246901	ATP-depende[illegible] Candida albicans]	46	28	219
61	4	3941	7930	gi|298032	EF [Strepto[illegible]	46	35	3990
132	4	5028	4093	gi|1511057	hypothetic [illegible] 869 [Methanococcus jannaschii]	46	25	936
170	4	4719	3652	pir|S51910|S519	G4 protein [illegible] a tarentolae	46	26	1068
191	7	9543	8284	gi|1041334	F54D5.7 [Ca[illegible] gans]	46	25	1260
253	1	1	396	gi|1204449	dihydrolip[illegible] sterase [Haemophilus influenzae]	46	35	396
264	3	437	973	gi|180189	cerebellar [illegible] ated antigen (CDR34) [Homo sapiens] gi|182737 cerebellar [illegible] related protein [Homo sapiens] pir|A29770 [illegible] degeneration-related protein - human	46	29	537
273	1	485	285	gi|607573	envelope gl[illegible] region [Human immunodeficiency virus type 1	46	35	201
350	1	3	563	gi|537052	ORF_f286 [E[illegible]	46	35	561
384	1	2	862	gi|1221884	(urea?) am[illegible] ilus influenzae]	46	31	861
410	4	1876	2490	gi|1110518	proton anti[illegible] mp [Mycobacterium smegmatis]	46	24	615
432	1	2663	1455	gi|1197634	orf4; puta[illegible] Method: conceptual translation supplied by author [M[illegible] gmatis]	46	27	1209
458	1	2419	1211	gi|15470	portal pr[illegible] SPP1]	46	30	1209
517	5	2477	4192	gi|1523812	orf5 [Bac[illegible]	46	23	1716
540	3	1512	1285	gi|215635	pacA [Bac[illegible]	46	30	228
587	2	649	1242	gi|537148	ORF_f181 [illegible]	46	29	594
1218	1	747	391	gi|1205456	single-str[illegible] xonuclease [Haemophilus influenzae]	46	30	357

TABLE 2

S. aureus - Putative coding ... proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene ...	...	% sim	% ident	length (nt)
3685	1	1	402	gi|450688	hsdM gene of protein - Es Escherichia	...ocol] pir|S38437|S38437 hsdM ...ir|S09629|S09629 hypothetical protein A - ...	46	33	402
4176	1	673	338	gi|351460	FIM-C.1 gene	...laevis]	46	31	336
37	7	4813	5922	gi|606064	ORF_f408 (Es		45	24	1110
38	16	11699	12004	gi|452192	protein tyrosi	(PTP-BAS, type 2) [Homo sapiens]	45	24	306
87	2	1748	2407	gi|1064813	homologous to	Bacillus subtilis]	45	23	660
101	12	14182	13385	gi|1001307	hypothetical	...ocystis sp.]	45	22	798
112	14	14791	13811	gi|1204389	H. influenza	...ng region HI0131 [Haemophilus influenzae]	45	23	981
145	4	4483	3461	gi|220578	open reading	...ilus]	45	20	1023
170	6	6329	4965	gi|238657	AppC=cytoch... eptide, 51	...subunit I homolog [Escherichia coli, K12.	45	27	1365
206	2	5230	4346	gi|1222056	aminotransfe...	...influenzae]	45	27	885
228	1	60	716	gi|160299	glutamic acid glutamic ac	...plasmodium falciparum] pir|A54514|A54514 ...precursor - Plasmodium alciparum	45	23	657
288	1	2	1015	gi|1255425	C13GR.2 gene	...abditis elegans]	45	23	1014
313	3	4339	3128	gi|581140	NADH dehydro	...ia coli]	45	30	1212
332	1	914	459	gi|870966	F47A4.2 [Cae	...sis]	45	20	456
344	1	3	221	gi|171225	kinesin-rela	...haromyces cerevisiae]	45	26	219
441	2	1501	1073	gi|142863	replication replication	...s (Bacillus subtilis] pir|B26580|B26580 ...tein - Bacillus ubtilis	45	27	429
672	1	2	982	gi|1511334	M. jannaschi	...A region MJ1323 [Methanococcus jannaschii]	45	22	981
763	3	1345	851	gi|606180	ORF_f310 (E		45	24	495
886	3	379	846	gi|726426	similar to [Caenorhat	elegans proteins F37C12.8 and 37C12.5	45	30	468
948	1	3	473	gi|156400	...yosin heav pir|A93958 sp|P02566|M	...(P-4)) [Caenorhabditis elegans] ...chain B - Caenorhabditis elegans ...MY CHAIN B (MHC B).	45	25	471
1158	1	2	376	gi|441155	transmission	...falciparum) [Plasmodium falciparum]	45	35	375
2551	3	4	285	gi|1276705	ORF287 gene	...rhodopurea]	45	28	282
3967	1	42	374	gi|976025	HrsA [Esche		45	28	333

TABLE 2

S. aureus - Putative coding

proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene		% sim	% ident	length (nt)
52	7	6931	5846	gi|467378	unknown (Bac		44	22	1086
138	8	6475	6849	gi,173028	thioredoxin	cerevisiae]	44	28	375
221	5	7032	5617	gi|153490	tetracenomyc	l export protein (Streptomyces laucescens]	44	21	1416
252	2	1331	1122	gi|1204989	hypothetical	22_9) (Haemophilus influenzae]	44	30	210
263	2	3265	2093	gi|1136221	carboxypepti	olfataricus]	44	26	1173
365	4	4963	3524	gi|1296822	orf1 gene pro	lus helveticus]	44	31	1440
543	3	1315	1833	gi|...063250	low homology acetyltrans	Bacillus lichiniformis and bleomycin myces verticillus (Bacillus subtilis]	44	24	519
544	4	3942	4892	gi|951460	FIM-C.1 gene	laevis]	44	32	951
792	1	1224	613	gi|205680	high molecul	ment (Rattus norvegicus]	44	28	612
44	18	11303	11911	gi|1511614	molybdopter jannaschii]	otide biosynthesis protein A (Methanococcus	43	27	609
59	8	3665	5128	gi|153490	tetracenomyc	l export protein (Streptomyces laucescens]	43	21	1464
59	10	5536	7527	gi|153022	lipase (Ste	idis]	43	22	1992
99	1	1346	681	gi|1419051	unknown (My	losis]	43	21	666
310	8	9402	12134	gi|397526	clumping fa	s aureus]	43	21	2733
432	3	2782	2303	pir|A60540|A605	sporozoite	- Plasmodium yoelii (fragment]	43	29	480
519	3	2547	3122	sp|Q06530|DHSU_	SULFIDE DEH 1.8.2.-)	YTOCHROME C) FLAVOPROTEIN CHAIN PRECURSOR (EC	43	23	576
4	13	12053	13321	gi|295671	selected as dependant	of a mutant of the subunit AC40 of DNA I III (Saccharomyces cerevisiae]	42	18	1269
94	2	1768	1091	gi|501027	ORF2 (Trypa		42	31	678
127	4	5791	4550	gi|42029	ORF1 gene	coli]	42	21	1242
297	3	1515	1036	gi|142790	ORF1; putat	s]	42	25	480
344	6	4097	3525	gi|40320	ORF 2 (AA	uringiensis]	42	30	573
512	1	2167	1115	gi|405957	yeeF (Esche		42	23	1053
631	1	2434	1223	gi|580920	rodD (gtaA) probable POLY(GLYC ACID BIOS	971) (Bacillus subtilis] pir|S06048|S06048 us subtilis sp|P13484|TAGE_BACSU PROBABLE -GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHOIC	42	24	1212

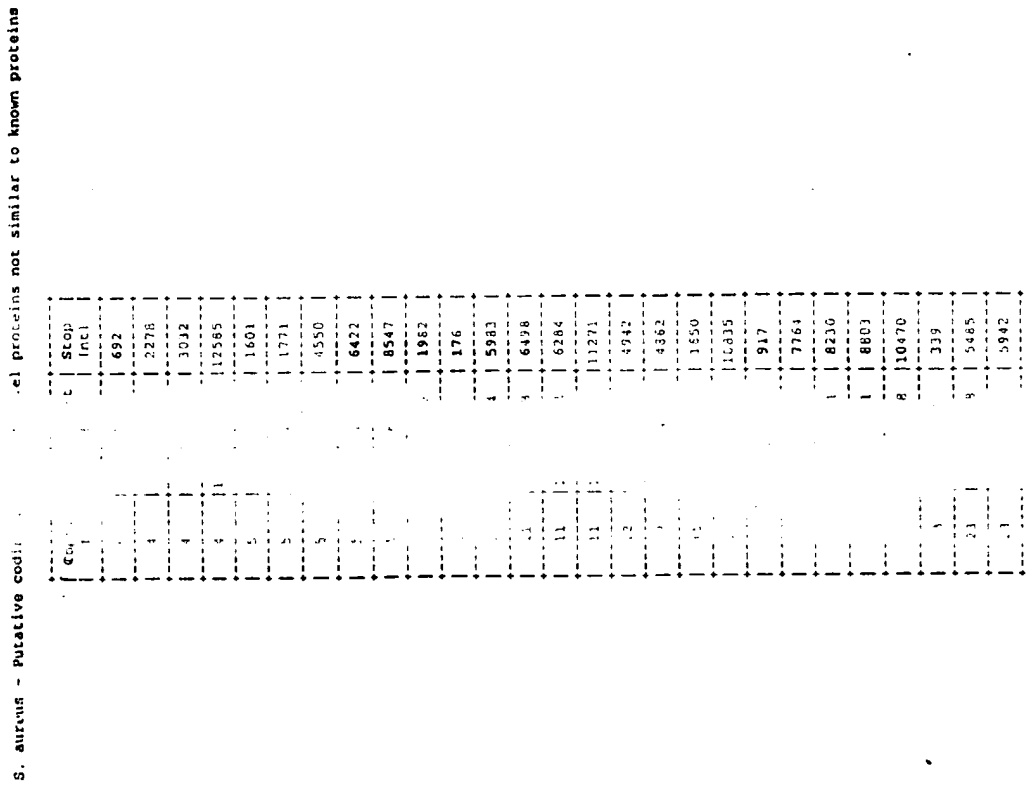
TABLE 2

S. aureus - Putative coding regio ... proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene ame		% sim	% ident	length (nt)
685	3	2359	1739	gi|1303784	YqeD [Baci ...		42	19	621
4132	1	787	395	gi|1022910	protein tyro ...	dictyostelium discoideum]	42	25	393
86	2	1375	884	gi|309506	spermidine/ ... spermidine/ ...	transferase [Mus saxicola] pir|S43430|S43430 transferase - spiny ouse (Mus saxicola)	41	30	492
191	12	14797	14075	gi|1124957	orf4 gene pr ...	cina barkeri]	41	22	723
212	6	2150	3127	gi|15873	observed 35. ...	bacteriophage 15]	41	26	978
213	3	1263	2000	gi|633692	TrsA [Yersini ...	]	41	18	738
408	4	2625	3386	gi|1197634	orf4; putativ ... author [Myc ...	Method: conceptual translation supplied by tis]	41	24	762
542	1	3	1103	gi|457146	rhoptry protein [...	coelii]	41	21	1101
924	1	2	475	pir|JH0148|JH01	nucleolin - rat		41	30	474
1562	1	1	402	gi|552184	asparagine-rich an ... asparagine rich a ...	-2 [Plasmodium falciparum] pir|S27826|S27826 5-2 - Plasmodium alciparum (fragment)	40	20	402
2395	1	518	261	pir|S42251|S422	hypothetica ...	pox virus	40	18	258
4077	1	3	305	gi|1055055	coded for b ... yk5c9.5; ... of F52C9. ...	c37g1.5; coded for by C. elegans cDNA ...na cDNA yk1a9.5; alternatively spliced form ...egans]	39	21	303
958	1	1003	503	gi|1255425	C33G8.2 ge ...	editis elegans]	37	25	501
59	12	8294	10636	gi|535260	STARP antig ...	ichenowi]	36	24	2343
63	5	3550	8079	gi|298032	EF [Streptoc ...		36	19	4530
544	3	2507	3601	gi|1015903	ORF YJR151c ...	erevisiae]	35	22	1095
63	4	1949	3574	gi|552195	circumspor ... CIRCUMSPOROZOITE ...	modium falciparum] sp|P05691|CSP_PLAFL 5) (FRAGMENT).	32	27	1626

TABLE 2

TABLE 3



S. aureus - Putative codi... ...el proteins not similar to known proteins

	Stop Intl
	692
	2278
	3032
	12585
	1601
	1771
	4550
	6422
	8547
	1962
	176
	5983
	6498
	6284
	11271
	4942
	4362
	1550
	10835
	917
	7761
	8230
	8803
	10470
	339
	5465
	5942

TABLE 3

S. aureus - Putative cod... ...al proteins not similar to known proteins

ID	...	Start	Stop (int)
		351	6881
		718	12830
		456	4185
		5142	5241
		1844	2402
		5...	849
		...	1524
		...	3005
		7	1388
			8575
		31	8728
		18	9379
		997	10087
		15	1019
		26	5801
		75	7261
		71	7621
		918	2961
		5	980
			6868
		1	16371
			20864
		2	21264
			627
		35	404
		6	426
		74	2324

TABLE 3

S. aureus - Putative coding sequences encoding novel proteins not similar to known proteins

Seq. ID No.	Start (nt)	Stop (nt)
41	2184	3263
43	10587	10129
45	13724	13536
47	13596	13994
49	6675	6297
51	[illegible]	6520
53	[illegible]	10976
55	15012	15424
57	288	1079
59	[illegible]	7778
61	[illegible]	362
63	1162	1316
65	718	370
67	5920	2245
69	442	287
71	7705	6319
73	[illegible]	8709
75	[illegible]	326
77	854	786
79	[illegible]	261
81	[illegible]	1228
83	1770	1560
85	[illegible]	18712
87	6914	3521
89	136	5322
91	485	8553
93	1366	1509

TABLE 3

S. aureus - Putative codi... of novel proteins not similar to known proteins

Contig ID	Start (nt)	Stop (nt)
5	3126	2802
5.	[illegible]	3570
5	[illegible]	4563
5	[illegible]	8378
5	[illegible]	16403
6.	[illegible]	1521
62	[illegible]	5757
63	1	336
6.	[illegible]	1781
6	[illegible]	2610
6	[illegible]	3904
6	[illegible]	6955
6	[illegible]	[illegible]
7	[illegible]	[illegible]
7	[illegible]	[illegible]
7	[illegible]	[illegible]
7	[illegible]	1228
7	[illegible]	1791
8	[illegible]	[illegible]
8	[illegible]	8653
8	[illegible]	8781
8	[illegible]	1232
8	[illegible]	9366
8	[illegible]	[illegible]
8	[illegible]	[illegible]
8	[illegible]	[illegible]
9	[illegible]	[illegible]

TABLE 3

S. aureus - Putative coding ... of ... nal proteins not similar to known proteins

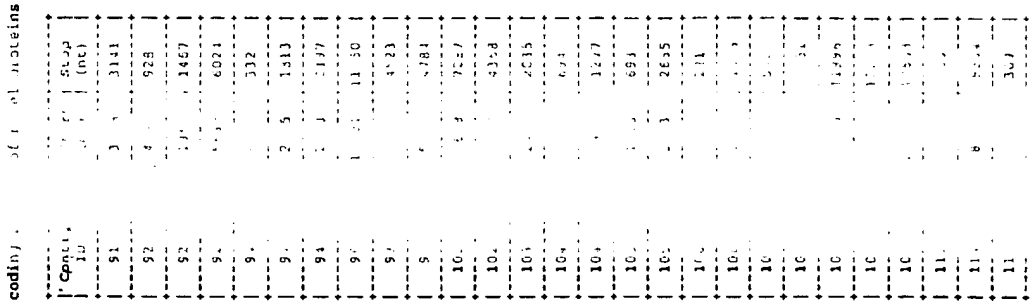


TABLE 3

S. aureus - Putative coding [illegible] proteins not similar to known proteins

Contig ID	St fo	[illegible]
116	62	[illegible]
116	[illegible]	[illegible]
116	103	[illegible]
120	37	3320
120	4	3869
120	95	9844
121	41	569
126	10	518
127	264	3396
127	40	[illegible]
131	67	[illegible]
135	71	[illegible]
134	[illegible]	[illegible]
135	5	[illegible]
135	[illegible]	[illegible]
138	1	[illegible]
138	6	6163
140	24	8032
140	2	[illegible]
140	12	[illegible]
142	13	[illegible]
142	18	[illegible]
142	[illegible]	[illegible]
141	[illegible]	[illegible]
141	[illegible]	[illegible]
146	[illegible]	[illegible]
147	5	[illegible]
149	[illegible]	[illegible]

TABLE 3

S. aureus - Putative coding [illegible] of [illegible] proteins not similar to known proteins

Contig ID		Stop (nt)
146	[illegible]	[illegible]
147	1	[illegible]
149	[illegible]	[illegible]
149	40	[illegible]
149	15	[illegible]
149	18	[illegible]
149	[illegible]	[illegible]
149	[illegible]	[illegible]
149	7	[illegible]
149	[illegible]	[illegible]
149	[illegible]	[illegible]
149	[illegible]	[illegible]
150	[illegible]	[illegible]
154	17	[illegible]
154	65	[illegible]
154	[illegible]	[illegible]
154	[illegible]	[illegible]
156	31	[illegible]
157	[illegible]	[illegible]
158	[illegible]	[illegible]
159	[illegible]	[illegible]
161	[illegible]	[illegible]
161	46	[illegible]
161	42	[illegible]
161	[illegible]	[illegible]
161	58	[illegible]
163	[illegible]	[illegible]

TABLE 3

S. aureus - Putative coding ... proteins not similar to known proteins

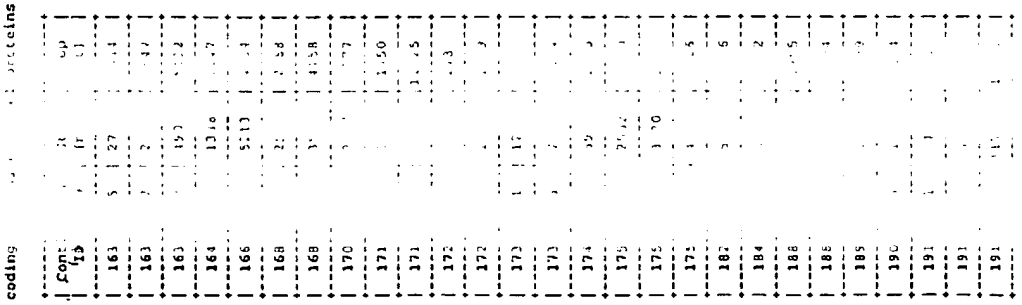


TABLE 3

S. aureus - Putative coding [illegible] not similar to known proteins

Contig Id	[illegible]
191	[illegible]
192	[illegible]
195	1902
195	2606
198	1591
201	[illegible]
203	116
206	[illegible]
206	[illegible]
206	[illegible]
212	[illegible]
212	[illegible]
212	[illegible]
212	[illegible]
213	70[illegible]
214	57[illegible]
214	[illegible]
214	[illegible]
214	[illegible]
217	[illegible]
218	[illegible]
218	[illegible]
220	[illegible]
220	[illegible]
220	[illegible]
220	[illegible]
220	[illegible]

TABLE 3

S. aureus - Putative coding [illegible] proteins not similar to known proteins

Contig SEQ ID	Cop (nt)
221	513
221	357
226	59
226	1359
226	1961
227	487
227	975
227	7121
227	2345
227	768
227	367
228	877
228	526
233	762
236	79
238	1191
239	405
241	1134
242	1363
243	76
244	547
244	962
245	258
246	15
246	733
249	712
250	49

TABLE 3

S. aureus - Putative coding [illegible] novel proteins not similar to known proteins

Contig ID	Start (nt)	Stop (nt)
254	1	156
256	35	244
257	1	227
260	1	580
261	1	96
261	2	81
264	15	39
264	25	33
264	1	107
267	1	11
268	5143	1700
272	862	416
272	1	19
272	9	69
272	0	35
276	1	901
278	2	1
278	1	18
278	8	27
285	5	6
288	1	50
288	6	276
289	1	55
290	2	12
291	2	2
291	5	51
295	1	19

TABLE 3

S. aureus - Putative coding f... ...l proteins not similar to known proteins

Contig ID	...tart (nt)	Stop (nt)
295	2728	2141
295	22..	2762
297	...68	365
298	2	205
300	2380	1928
301	2794	2624
304	3	194
306	109	654
306	103	4357
307	...	...139
307	...	3495
308	1	654
308	1200	799
308	...	2132
313	...	...519
314	10	702
316	982	1341
316	27	3165
317	2	1114
317	157	1458
321	...	...217
321	31	...130
321	...	...794
322	...	...43
326	45	412
326	...	...47
328	...	...

TABLE 3

S. aureus - Putative coding r... proteins not similar to known proteins

Contig ID	Stop (nt)
328	3276
329	719
329	1212
329	1833
330	259
330	1623
332	2294
332	5138
333	3128
335	413
337	516
340	156
341	241
341	592
341	11
341	598
344	269
345	58
346	2
350	3
352	265
352	4256
352	604
352	2497
352	1554
359	6
362	6

TABLE 3

S. aureus - Putative coding r... of ... proteins not similar to known proteins

Contig ID	St... (bp)	...
364	...	...
364	...	...
365	...	...
365	...	...
365	...	...
366	...	...
367	...	...
368	...	...
375	...	...
380	...	...
389	...	...
390	...	...
390	...	...
391	...	...
395	...	...
396	...	...
398	...	...
399	...	...
401	...	...
402	...	...
404	...	...
408	...	...
408	...	...
408	...	...
410	...	...
413	1	...
416	...	...

TABLE 3

S. aureus - Putative coding re... in ... proteins not similar to known proteins

Contig ID				
416				
416				
417				
417				
420				
422				
431				
432				
433				
433				
434				
434				
440				
442				
443				
444				
444				
451				
453				
453				
453				
453				
455				
455				
459				
462				
466				

TABLE 3

S. aureus - Putative coding re[illegible] of n[illegible]el pr[illegible]ins not similar to known proteins

Contig ID	[illegible] nt	[illegible]
467	[illegible]4	[illegible]
468	2	[illegible]
469	[illegible]48[illegible]	92[illegible]
469	238[illegible]	3372
469	346[illegible]	[illegible]
470	77	[illegible]
470	[illegible]09	[illegible]
470	[illegible]33[illegible]	[illegible]
470	735[illegible]	[illegible]
470	817	[illegible]
471	[illegible]	[illegible]
471	56	[illegible]
476	7[illegible]	[illegible]
477	2	[illegible]
477	[illegible]76[illegible]	[illegible]
477	206	[illegible]
480	401[illegible]	[illegible]
481	95[illegible]	[illegible]
486	51[illegible]	[illegible]
487	179	[illegible]
488	71[illegible]	[illegible]
492	[illegible]	[illegible]
493	[illegible]	[illegible]
493	[illegible]	[illegible]
502	[illegible]	[illegible]
504	90	[illegible]
505	15[illegible]	[illegible]

TABLE 3

S. aureus - Putative coding regions from known proteins not similar to known proteins

Contig ID		Score	
511		1741	
512		583	
515		609	
517		2179	
520	4	2097	
520	6	90	
527	1	1	
528	1	337	335
529		167	1104
530		529	534
536		208	352
538		1364	
538		2233	
538		3931	
538		1446	
540		396	
540	2	1595	
541	1	1	
541	2	13	
542		931	1272
545		0	734
551		2117	1129
555		3	21
558		1107	1
558		2044	
558		210	
558		2169	

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start	Stop	Length
558	7	[illegible]	[illegible]	[illegible]
558	8	305	[illegible]	[illegible]
558	9	[illegible]	[illegible]	[illegible]
560	[illegible]	[illegible]	[illegible]	[illegible]
565	3	[illegible]	[illegible]	[illegible]
571	1	[illegible]	[illegible]	[illegible]
571	3	[illegible]	[illegible]	[illegible]
577	1	[illegible]	[illegible]	[illegible]
577	2	[illegible]	[illegible]	[illegible]
579	1	[illegible]	[illegible]	[illegible]
579	2	[illegible]	[illegible]	[illegible]
583	1	[illegible]	[illegible]	[illegible]
585	1	[illegible]	[illegible]	[illegible]
587	1	2	[illegible]	[illegible]
588	[illegible]	[illegible]	[illegible]	2
588	[illegible]	[illegible]	[illegible]	[illegible]
590	3	[illegible]	[illegible]	[illegible]
592	2	[illegible]	[illegible]	[illegible]
593	1	[illegible]	[illegible]	[illegible]
593	2	[illegible]	[illegible]	[illegible]
595	1	[illegible]	[illegible]	[illegible]
596	3	[illegible]	[illegible]	[illegible]
602	1	[illegible]	[illegible]	[illegible]
603	[illegible]	[illegible]	[illegible]	[illegible]
606	1	32	[illegible]	[illegible]
607	[illegible]	[illegible]	[illegible]	[illegible]
610	[illegible]	[illegible]	[illegible]	[illegible]

TABLE 3

S. aureus - Putative coding regions of no proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
612	1	3	500
616	1	691	956
617	1	736	891
622	1	36	347
625	1	204	554
627	1	67	411
628	1	901	1
631	1	478	1
634	1	1448	953
636	1	189	565
636	2	1929	1064
637	2	2323	1994
638	1	227	1061
639	1	518	261
639	1	137	1
641	1	118	1
642	1	161	1
642	1	224	1
643	1	3	1
645	1	153	1
645	1	2021	1
645	1	2940	1
648	1	2	1064
660	1	77	601
660	2	576	724
661	1	172	1
664	1	88	1

TABLE 3

S. aureus - Putative coding regions not similar to known proteins

Contig ID	Start (nt)	Stop (nt)
667	3	41
668	1	33
671	812	51[illegible]
673	3	33[illegible]
674	865	[illegible]
679	1	[illegible]
679	1586	[illegible]
688	1251	[illegible]
688	1350	[illegible]
694	3	14[illegible]
696	818	432
706	[illegible]	42[illegible]
709	1183	[illegible]
711	3	9[illegible]
715	3	16
716	2	[illegible]
721	133	[illegible]
722	763	[illegible]
723	165	[illegible]
723	147	[illegible]
727	2	[illegible]
729	266	141
731	130	32[illegible]
735	2	21[illegible]
736	3	78
738	2	26
742	3	2[illegible]

TABLE 3

S. aureus - Putative coding reg of nove[illegible] eins not similar to known proteins

Contig ID	Start (nt)
745	114
748	282
749	68[illegible]
751	90[illegible]
755	9[illegible]
755	520
758	663
764	1033
767	1
768	2
771	1058
778	175
785	177
787	12[illegible]
791	3
799	1[illegible]
804	30
805	3
808	219
810	2221
810	1774
812	38
817	714
818	48
819	15
819	17
820	19

TABLE 3

S. aureus - Putative coding regio[illegible] novel [illegible]s not similar to known proteins

Contig ID	Start (nt)
828	506
829	48
830	578
832	594
835	320
840	491
845	912
850	203
853	715
860	2
864	18
864	383
864	1676
870	1
873	906
875	584
877	1661
878	981
879	1567
881	1
882	389
890	2
905	793
906	452
912	173
913	3
913	1092

TABLE 3

S. aureus - Putative coding regions not similar to known proteins

Contig ID	Start (nt)
915	6
915	169
921	126
927	1578
928	2
929	2
932	2
934	1
936	105
937	2
945	220
945	649
946	1702
949	1
951	3
955	3
960	723
963	1
965	691
966	105
969	3
971	12
974	319
976	692
977	2
982	1926
984	589

TABLE 3

S. aureus - Putative coding regions not similar to known proteins

Contig ID	Start	Int
987		3
993	1	1
994		920
1004		557
1014		624
1015		2
1016		2[illegible]
1019		[illegible]
1022		83
1024		59
1024		27[illegible]
1030		673
1032		355
1040	1	794
1043		3
1044		115
1047		1
1051		7[illegible]
1051		[illegible]
1063		[illegible]
1069		[illegible]
1069		7[illegible]
1075		76
1077		9[illegible]
1081		58
1086	1	1
1087		2[illegible]6

TABLE 3

S. aureus - Putative coding regions not similar to known proteins

Contig ID	
1088	
1096	
1098	
1100	
1100	
1101	
1102	
1107	
1114	
1115	
1119	
1129	
1132	
1133	
1144	
1147	
1153	
1154	
1159	
1161	
1164	
1171	
1171	
1183	
1195	
1196	
1200	

TABLE 3

S. aureus - Putative coding re[illegible] not similar to known proteins

Contig ID
1203
1222
1232
1240
1247
1271
1286
1295
1306
1314
1316
1359
1370
1371
1374
1378
1392
1411
1433
1450
1453
1471
1477
1502
1518
1534
1546

TABLE 3

S. aureus - Putative coding regions not similar to known proteins

Contig ID
1547
1583
1587
1602
1629
1665
1760
1762
1876
1895
1931
1976
2055
2056
2150
2157
2164
2175
2212
2338
2342
2352
2352
2355
2356
2359
2421

TABLE 3

S. aureus - Putative coding regions s t similar to known proteins

Contig ID	ORF ID
3046	1
3049	1
3050	1
3052	1
3065	1
3070	1
3075	1
3080	1
3092	1
3093	1
3100	1
3103	1
3118	1
3123	1
3127	1
3138	1
3142	1
3144	1
3151	1
3155	2
3168	1
3205	1
3282	1
3303	2
3371	2
3558	1
3558	2

TABLE 3

S. aureus - Putative coding reg[illegible] [illegible]ins not similar to known proteins

Contig ID	[illegible] (nt)
3046	[illegible]
3049	[illegible]
3050	[illegible]
3052	[illegible]
3065	[illegible]
3070	[illegible]
3075	[illegible]
3080	[illegible]
3092	[illegible]
3093	[illegible]
3100	[illegible]
3103	[illegible]
3118	[illegible]
3123	[illegible]
3127	[illegible]
3138	[illegible]
3142	[illegible]
3144	[illegible]
3151	[illegible]
3155	[illegible]
3168	[illegible]
3205	[illegible]
3282	[illegible]
3303	[illegible]
3371	[illegible]
3558	[illegible]
3558	[illegible]

TABLE 3

S. aureus - Putative coding regions encoding novel proteins not similar to known proteins

Contig ID	Stop (nt)
3568	377
3595	380
3618	238
3618	402
3622	358
3622	398
3642	439
3649	398
3651	314
3664	637
3674	402
3677	311
3704	402
3726	269
3765	256
3779	357
3794	135
3794	377
3796	375
3801	262
3806	298
3807	389
3815	400
3827	320
3842	392
3853	399
3855	324

TABLE 3

S. aureus - Putative coding [illegible] novel proteins not similar to known proteins

Contig ID	Start (nt)	Stop (nt)
3857	[illegible]	235
3861	[illegible]	297
3865	[illegible]	399
3897	[illegible]	173
3897	[illegible]	400
3898	[illegible]	401
3921	[illegible]	342
3927	[illegible]	375
3930	[illegible]	234
3946	[illegible]	382
3951	[illegible]	377
3965	[illegible]	344
3973	[illegible]	400
3981	[illegible]	311
3998	[illegible]	356
4001	[illegible]	296
4003	[illegible]	335
4018	[illegible]	259
4018	[illegible]	401
4021	[illegible]	345
4043	[illegible]	344
4054	[illegible]	344
4060	[illegible]	150
4070	[illegible]	324
4072	[illegible]	390
4073	[illegible]	285
4077	[illegible]	372

TABLE 3

S. aureus - Putative coding re... to known proteins not similar to known proteins

Contig # ID	Start (nt)	Stop (nt)
4083		359
4090		368
4101		297
4105		306
4107		286
4119		339
4121		372
4123		230
4127		341
4128		331
4130		415
4146		381
4157		255
4186		251
4224		255
4239		345
4242		355
4252		236
4253		174
4256		323
4258		339
4267		144
4271		304
4287		163
4289		311
4302		305
4304		155

TABLE 3

S. aureus - Putative coding regions not similar to known proteins

Contig ID	Start (nt)	Stop (nt)
4304	5	314
4306		151
4318	5	289
4322		148
4331		221
4331	5	361
4338	3	394
4346	171	277
4367	117	311
4373	2	
4381	4	325
4384	3	309
4397		311
4402		249
4403		
4406		
4411		
4411		
4412		
4416		23
4424		
4443		
4471	5	32
4478		271
4482	5	285
4485	1	364
4491		

TABLE 3

S. aureus - Putative coding r... s of [illegible] ...eins not similar to known proteins

Contig ID	[illegible]
4495	3
4496	[illegible]
4500	[illegible]
4511	[illegible]
4518	[illegible]
4526	[illegible]
4527	[illegible]
4532	[illegible]
4542	[illegible]
4567	[illegible]
4573	[illegible]
4578	[illegible]
4619	[illegible]
4620	[illegible]
4662	[illegible]
4669	[illegible]
4680	[illegible]
4690	[illegible]

Table 4

ORF	SEQ ID NO	BLAST HOMOLOG	Antigenic Regions Region 1	Region 2	Region 3	Region 4
168_6	5192	lipoprotein	36-45	84-103	152-161	176-185
238_1	5193	chrA	21-39	48-58	84-95	232-249
51_2	5194	OppB gene product (B. sub	20-36	70-79	100-112	121-131
278_3	5195	lipoprotein 1	20-29	59-73	85-97	162-171
276_2	5196	lipoprotein	21-33	65-74	177-186	211-220
45_4	5197	ProX	28-37	59-69	85-100	120-129
316_8	5198	hypothetical protein	45-54	88-97	182-192	243-253
154_15	5199	unknown	31-40	48-58	79-88	95-104
228_3	5200	unknown	25-38	40-52	64-74	80-89
228_6	5201	unknown	29-41	89-101	128-143	173-184
50_1	5202	unknown	21-33	52-61	168-182	197-206
112_7	5203	iron-binding periplasmic	21-31	58-67	92-101	111-120
442_1	5204	unknown	30-39	91-100	122-137	182-192
66_2	5205	unknown	50-59	104-116	127-136	167-182
304_2	5206	Q-binding periplasmic	19-28	48-57	75-84	103-116
44_1	5207	hypothetical protein	27-36	86-95	129-138	192-201
161_4	5208	SphX	27-44	149-161	166-175	201-210
46_5	5209	cmpC (permease)	21-33	61-70	83-92	100-109
942_1	5210	traH [Plasmid pSK41]	83-92	109-118	127-142	
5_4	5211	ORF (S. aureus)	12-22	87-96	111-120	151-160
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
771_1	5215	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
999_1	5216	[illegible] trithorax prot. (D	7-15	[illegible]	157-166	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
288_2	5219	cell wall enzyme	14-23	89-98		114-12
596_2	5220	penicillin binding protein 2b	40-49	59-68	76-87	106-115
217_5	5221	fibronectin/fibrinogen bindi	28-37	40-49	62-71	93-111
217_6	5222	fibronectin/fibrinogen bp	10-19	31-40	54-62	73-92
528_3	5223	myosin cross reactive prote	4-13	29-47	60-73	90-99
171_11	5224	EF	20-31	91-110		
63_4	5225	penicillin binding protein 2b	12-21	59-68	95-104	
353_2	5226		46-55	62-71		
743_1	5227	29 kDa protein in fimA regi	23-32	68-79	94-103	175-184
342_4	5228	Twitching motility	10-19	48-60	83-92	111-121
69_3	5229	arabinogalactan protein	97-106	132-141	158-167	180-189
70_6	5230	nodulin	36-45	48-57	137-160	179-188
129_2	5231	glycerol diester phosphodie	8-17	41-50	55-74	97-106
58_5	5232	PBP (S. aureus)	26-35	70-79	117-126	152-161
188_3	5233	MHC class II analog (S. aure	72-81	94-103	115-124	136-145
236_6	5234	histidine kinase domain (Dic	24-33	52-67	81-94	106-121
310_8	5235	clumping factor (S. aureus)	59-71	77-86	93-102	118-127
601_1	5236	novel antigen/ORF2 (S. aur	45-54	91-104	108-117	186-195
544_3	5237	ORF YJR151c (S. cerevisae)	76-90	101-111	131-140	154-164
662_1	5238	MHC class II analog (S. aure	22-32	71-80	89-98	114-122
87_7	5239	5' nucleotidase precursor ('	29-45	62-71	105-114	125-137
120_1	5240	B65G gene product (B. sub	102-111			

Table 4

ORF	Antigenic Regions (cont)					
	Region 5	Region 6	Region 7	Region 8	Region 9	Region 10
168_6	244-272	303-315				
238_1	260-269	291-301	308-317			
51_2	140-152	188-208	211-220	256-266	273-283	
278_3	198-209					
276_2	255-268					
45_4	177-199	221-230	234-243	268-279	284-293	304-313
316_8						
154_15	148-157	177-187	202-211			
228_3	101-119	139-154	166-181			
228_6						
50_1						
112_7	136-149	197-211	218-229	253-273		
442_1	199-210	247-257	264-277	287-309		
66_2						
304_2	178-187	250-259				
44_1						
161_4						
46_5	131-141	162-176	206-215	243-252	264-273	285-294
942_1						
20_4	[illegible]	217-234	[illegible]	4-336	366-3	
328_2						
520_2						
[illegible]						
899_1						
853_1						
287_1	154-164					
288_2						
596_2	121-130					
217_5	244-253	259-268	288-297	302-311		
217_6	144-158	174-183	188-197	207-216	226-242	
528_3						
171_11						
63_4						
353_2						
743_1	197-207					
342_4						
69_3	195-211					
70_6	206-215	263-272	291-301	331-340	358-371	390-414
129_2	117-127	141-157	168-183	202-211	222-231	261-270
58_5	184-203	260-269	275-299	330-344	372-381	424-433
188_3						
236_6	138-147	163-172	187-198	244-261	268-278	308-317
310_8	131-140	144-153	177-186	190-199	204-213	216-227
601_1	208-218					
544_3	170-179	184-193	224-235	274-287	327-336	352-361
662_1						
87_7						
120_1						

Table 4

ORF	Antigenic Regions (cont)					
	Region 11	Region 12	Region 13	Region 14	Region 15	Region 16
168_6						
238_1						
51_2						
278_3						
276_2						
45_4						
316_8						
154_15						
228_3						
228_6						
50_1						
112_7						
442_1						
66_2						
304_2						
44_1						
161_4						
46_5	306-315					
328_2						
520_2						
999_1						
853_1						
287_1						
288_2						
596_2						
217_5						
217_6						
528_3						
171_11						
63_4						
353_2						
743_1						
342_4						
69_3						
70_6	453-471	506-515				
129_2	296-315					
58_5						
188_3						
236_6	358-377	410-423	428-439	442-457	467-476	480-493
310_8	238-251	256-275	281-290	296-310	314-333	338-347
601_1						
544_3						
662_1						
87_7						
120_1						

Table 4

ORF	Antigenic Regions (cont)					
	Region 17	Region 18	Region 19	Region 20	Region 21	Region 22
168_6						
238_1						
51_2						
278_3						
276_2						
45_4						
316_8						
154_15						
228_3						
228_6						
50_1						
112_7						
442_1						
66_2						
304_2						
44_1						
161_4						
46_5						
20_4						
328_2						
520_2						
771_1						
949_1						
853_1						
287_1						
288_2						
596_2						
217_5						
217_6						
528_3						
171_11						
63_4						
353_2						
743_1						
342_4						
69_3						
70_6						
129_2						
58_5						
188_3						
236_6						
310_8	357-366	370-379	429-438	443-452	478-487	551-560
601_1						
544_3						
662_1						
87_7						
120_1						

Table 4

ORF	Antigenic Regions (cont)					
	Region 23	Region 24	Region 25	Region 26	Region 27	Region 28
168_6						
238_1						
51_2						
278_3						
276_2						
45_4						
316_8						
154_15						
228_3						
228_6						
50_1						
112_7						
442_1						
66_2						
304_2						
44_1						
161_4						
46_5						
[illegible]						
20_4						
328_2						
520_2						
771_1						
853_1						
287_1						
288_2						
596_2						
217_5						
217_6						
528_3						
171_11						
63_4						
353_2						
743_1						
342_4						
69_3						
70_6						
129_2						
58_5						
188_3						
236_6						
310_8	622-632	670-685	708-718	823-836	858-867	877-886
601_1						
544_3						
662_1						
87_7						
120_1						

Table 4

ORF	Antigenic Regions (cont)	
	Region 29	Region 30
168_6		
238_1		
51_2		
278_3		
276_2		
45_4		
316_8		
154_15		
228_3		
228_6		
50_1		
112_7		
442_1		
66_2		
304_2		
44_1		
161_4		
46_5		
94[illegible]		
[illegible]		
20_4		
328_2		
520_2		
771_1		
939_1		
853_1		
287_1		
288_2		
596_2		
217_5		
217_6		
528_3		
171_11		
63_4		
353_2		
743_1		
342_4		
69_3		
70_6		
129_2		
58_5		
188_3		
236_6		
310_8		
601_1		
544_3		
662_1		
87_7		
120_1		

Table 4

ORF		BLAST	Antigenic Regions			
		HOMOLOG	Region 1	Region 2	Region 3	Region 4
46_1	5241	aldehyde dehydrogenase	8-17	36-52	83-96	112-121
63_4	5242	glycerol ester hydrolase (P.	9-26	57-73	93-107	123-133
174_6	5243	ketopantoate hydroxymeth	71-80	203-212	242-254	265-274
206_16	5244	ornithine acetyltransferase	1-10	34-43	54-63	194-210
267_1	5245	NaH-antiporter protein (E. h	120-129	332-347	398-408	
322_1	5246	acriflavin resistance protein	58-75	153-164	203-231	264-284
415_2	5247	transport ATP-binding prot(	108-126	218-227	298-308	315-334
214_3	5248	2-nitropropane dioxygenase	123-136	216-233	283-292	297-306
587_3	5249	clumping factor	5-14	43-54	59-68	76-95
685_1	5250	signal peptidase	59-68	72-81	86-95	99-108
54_3	5251	fibronectin binding protein I	23-32	37-46	50-59	89-98
54_4	5252	fibronectin binding protein I	43-52	66-75	95-104	147-156
54_5	5253	fibronectin binding protein I	49-60	81-90		
54_6	5254	fibronectin binding protein I	55-71	82-97	139-158	175-186
328_1	5255	lipoprotein (H. flu)	11-20	61-70	96-105	

Table 4

ORF	Antigenic Regions (cont)					
	Region 5	Region 6	Region 7	Region 8	Region 9	Region 10
46_1	215-242	333-352	376-385	416-432	[illegible]	
63_4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
174_6						
206_16	239-259	[illegible]				
267_1						
322_1	298-319	350-359				
415_2	344-353	371-380	395-404	456-465	486-495	518-527
214_3	318-337	365-375				
587_3	106-115	142-151	156-166	173-182	186-198	204-213
685_1	113-122	130-145				
54_3	128-138	185-194	217-226	251-260	268-277	295-305
54_4	175-188	191-200	203-212	220-229		
54_5						
54_6	220-230	287-304	317-326	344-353	364-373	378-387
328_1						

Table 4

ORF	Antigenic Regions (cont)					
	Region 11	Region 12	Region 13	Region 14	Region 15	Region 17
46_1						
63_4	306-315	319-328	366-376	395-420	453-462	467-476
174_6						
206_16						
267_1						
322_1						
415_2	539-555					
214_3						
587_3	217-226	278-287	318-327	332-342	351-360	377-386
685_1						
54_3	316-325	329-345	355-372	387-396	416-425	438-448
54_4						
54_5						
54_6	396-407	427-436	514-531	541-550	569-578	612-622
328_1						

Table 4

ORF	Antigenic Regions (cont)					
	Region 18	Region 19	Region 20	Region 21	Region 22	Region 23
46_1						
[illegible]	[illegible] 500	513-[illegible]				
174_6						
206_16						
267_1						
322_1						
415_2						
214_3						
587_3	396-405	426-442	459-470	485-494	505-514	531-562
685_1						
54_3	455-462	472-491	517-536			
54_4						
54_5						
54_6	639-648	673-681	703-715	723-732	749-760	772-788
328_1						

Table 4

ORF	Antigenic Regions (cont)					
	Region 24	Region 25	Region 26	Region 27	Region 28	Region 29
46_1						
63_4						
174_6						
206_16						
267_1						
322_1						
415_2						
214_3						
587_3	567-578	584-601	607-840	844-854	858-870	877-886
685_1						
54_3						
54_4						
54_5						
54_6	793-802	811-826	834-848	866-876	893-903	907-918
328_1						

Table 4

ORF	Antigenic Regions (cont)		
	Region 30	Region 31	
46_1			
63_4			
174_6			
206_16			
267_1			
322_1			
415_2			
214_3			
587_3	889-911	927-936	
685_1			
54_3			
54_4			
54_5			
54_6	925-944	951-997	
328_1			